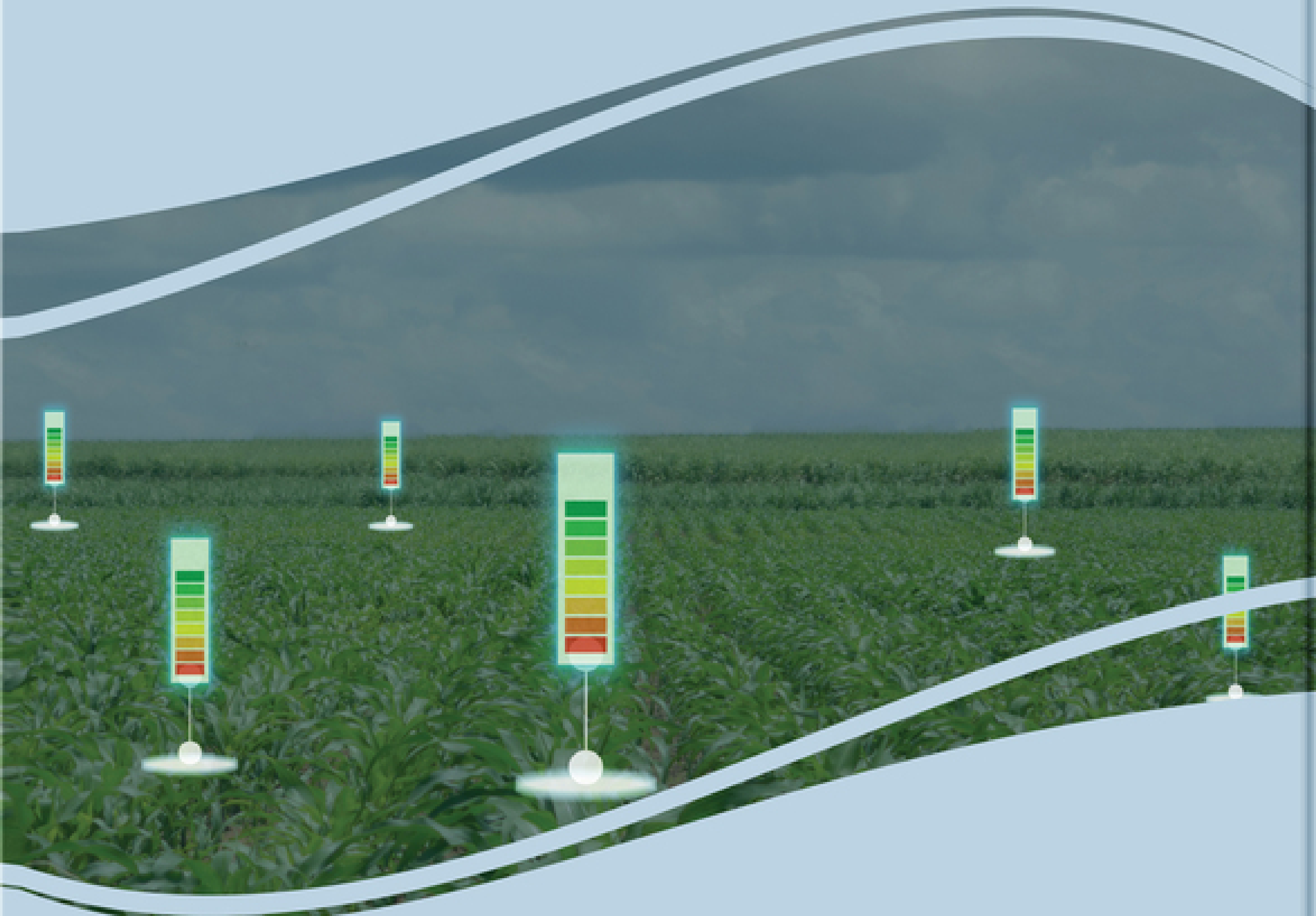
Premier Reference Source

Modern Techniques for Agricultural Disease Management and Crop Yield Prediction



N. Pradeep, Sandeep Kautish, C.R. Nirmala, Vishal Goyal, and Sonia Abdellatif

IGI Global
DISSEMINATOR OF KNOWLEDGE

Modern Techniques for Agricultural Disease Management and Crop Yield Prediction

N. Pradeep
Bapuji Institute of Engineering and Technology, India

Sandeep Kautish
Lord Buddha Education Foundation, Nepal & Asia Pacific University, Malaysia

C.R. Nirmala
Bapuji Institute of Engineering and Technology, India

Vishal Goyal
Punjabi University, Patiala, India

Sonia Abdellatif
Seoul National University, South Korea

A volume in the Advances in Environmental Engineering and Green Technologies (AEEGT) Book Series

IGI Global
DISSEMINATOR OF KNOWLEDGE

Published in the United States of America by
IGI Global
Engineering Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA, USA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

Library of Congress Cataloging-in-Publication Data

Names: Pradeep, N., 1977- editor.
Title: Modern techniques for agricultural disease management and crop yield prediction / N. Pradeep, Sandeep Kautish, C.R. Nirmala, Vishal Goyal, and Sonia Abdellatif, editors.
Description: Hershey, PA : Engineering Science Reference, 2019.
Identifiers: LCCN 2019005482| ISBN 9781522596325 (hardcover) | ISBN 9781522596332 (softcover) | ISBN 9781522596349 (ebook)
Subjects: LCSH: Agriculture--Data processing. | Agricultural innovations. | Precision farming. | Phytopathogenic microorganisms--Control. | Crop yields.
Classification: LCC S494.5.D3 M63 2019 | DDC 338.10285--dc23 LC record available at https://lccn.loc.gov/2019005482

This book is published in the IGI Global book series Advances in Environmental Engineering and Green Technologies (AEEGT) (ISSN: 2326-9162; eISSN: 2326-9170)

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

All work contributed to this book is new, previously-unpublished material. The views expressed in this book are those of the authors, but not necessarily of the publisher.

For electronic access to this publication, please contact: eresources@igi-global.com.

Advances in Environmental Engineering and Green Technologies (AEEGT) Book Series

Sang-Bing Tsai
University of Electronic Science and Technology of China Zhongshan Institute, China
Ming-Lang Tseng
Lunghwa University of Science and Technology, Taiwan
Yuchi Wang
University of Electronic Science and Technology of China Zhongshan Institute, China

ISSN:2326-9162
EISSN:2326-9170

Mission

Growing awareness and an increased focus on environmental issues such as climate change, energy use, and loss of non-renewable resources have brought about a greater need for research that provides potential solutions to these problems. Research in environmental science and engineering continues to play a vital role in uncovering new opportunities for a "green" future.

The **Advances in Environmental Engineering and Green Technologies (AEEGT)** book series is a mouthpiece for research in all aspects of environmental science, earth science, and green initiatives. This series supports the ongoing research in this field through publishing books that discuss topics within environmental engineering or that deal with the interdisciplinary field of green technologies.

Coverage

- Pollution Management
- Renewable Energy
- Policies Involving Green Technologies and Environmental Engineering
- Green Technology
- Air Quality
- Electric Vehicles
- Radioactive Waste Treatment
- Industrial Waste Management and Minimization
- Water Supply and Treatment
- Alternative Power Sources

IGI Global is currently accepting manuscripts for publication within this series. To submit a proposal for a volume in this series, please contact our Acquisition Editors at Acquisitions@igi-global.com or visit: http://www.igi-global.com/publish/.

The Advances in Environmental Engineering and Green Technologies (AEEGT) Book Series (ISSN 2326-9162) is published by IGI Global, 701 E. Chocolate Avenue, Hershey, PA 17033-1240, USA, www.igi-global.com. This series is composed of titles available for purchase individually; each title is edited to be contextually exclusive from any other title within the series. For pricing and ordering information please visit http://www.igi-global.com/book-series/advances-environmental-engineering-green-technologies/73679. Postmaster: Send all address changes to above address.

Titles in this Series

For a list of additional titles in this series, please visit: www.igi-global.com/book-series

Handbook of Research on Energy-Saving Technologies for Environmentally-Friendly Agricultural Development
Valeriy Kharchenko (Federal Scientific Agroengineering Center VIM, Russia) and Pandian Vasant (Universiti Teknologi Petronas, Malaysia)
Engineering Science Reference • copyright 2020 • 554pp • H/C (ISBN: 9781522594208) • US $295.00 (our price)

Hydrology and Water Resources Management in Arid, Semi-Arid, and Tropical Regions
Christopher Misati Ondieki (Kenyatta University, Kenya) and Johnson Utu Kitheka (South Eastern Kenya University, Kenya)
Engineering Science Reference • copyright 2019 • 356pp • H/C (ISBN: 9781799801634) • US $215.00 (our price)

Global Perspectives on Air Pollution Prevention and Control System Design
G. Venkatesan (Anna University, India) and Jaganthan Thirumal (Anna University, India)
Engineering Science Reference • copyright 2019 • 345pp • H/C (ISBN: 9781522572893) • US $195.00 (our price)

Retrofitting for Optimal Energy Performance
Adrian Tantau (Bucharest University of Economic Studies, Romania)
Engineering Science Reference • copyright 2019 • 339pp • H/C (ISBN: 9781522591047) • US $245.00 (our price)

Advanced Design of Wastewater Treatment Plants Emerging Research and Opportunities
Athar Hussain (Ch. Brahm Prakash Government Engineering College, India) and Ayushman Bhattacharya (Ch. Brahm Prakash Government Engineering College, India)
Engineering Science Reference • copyright 2019 • 350pp • H/C (ISBN: 9781522594413) • US $195.00 (our price)

Advanced Multi-Criteria Decision Making for Addressing Complex Sustainability Issues
Prasenjit Chatterjee (MCKV Institute of Engineering, India) Morteza Yazdani (Universidad Loyola Andalucía, Spain) Shankar Chakraborty (Jadavpur University, India) Dilbagh Panchal (Dr. B. R. Ambedkar National Institute of Technology (NIT) Jalandhar, India) and Siddhartha Bhattacharyya (RCC Institute of Information Technology Kolkata, India)
Engineering Science Reference • copyright 2019 • 360pp • H/C (ISBN: 9781522585794) • US $195.00 (our price)

Amelioration Technology for Soil Sustainability
Ashok K. Rathoure (Biohm Consultare Pvt Ltd, India)
Engineering Science Reference • copyright 2019 • 280pp • H/C (ISBN: 9781522579403) • US $185.00 (our price)

IGI Global
DISSEMINATOR OF KNOWLEDGE

701 East Chocolate Avenue, Hershey, PA 17033, USA
Tel: 717-533-8845 x100 • Fax: 717-533-8661
E-Mail: cust@igi-global.com • www.igi-global.com

Editorial Advisory Board

Table of Contents

Detailed Table of Contents

in recent days to find the leaf area and leaf count accurately. In this chapter, the recent techniques in image-based plant phenotyping and their limitations are discussed. Also, this chapter discusses a new plant segmentation method based on wavelet and leaf count methods based on Circular Hough Transform and deep learning model, which overcomes the drawbacks of recent methods. These methods are experimented with Computer Vision Problems in Plant Phenotyping (CVPPP) benchmark datasets.

Plants play one of the main roles in our ecosystem. Manual identification for the leaves sometimes leads to greater difference due to look alike. People often get confused with lookalike leaves which mostly end in loss of life. Authentication of original leaf with look-alike leaf is very essential nowadays. Disease identification of plants are proved to be beneficial for agro-industries, research, and eco-system balancing. In the era of industrialization, vegetation is shrinking. Early detection of diseases from the dataset of leaf can be rewarding and help in making our environment healthier and green. Implementation involves proper data acquisition where pre-processing of images is done for error correction if present in the raw dataset. It is followed by feature extraction stage to get the best results in further classification stage. K-mean, PCA, and ICA algorithms are used for identification and clustering of diseases in plants. The implementation proves that the proposed method shows promising result on the basis of histogram of gradient (HoG) features.

Smart farming is a development that highlights the use of technologies such as the internet of things, cloud computing, machine learning, and artificial intelligence in the farm management cycle. For sustainable agriculture to adapt the ongoing change in climate and social structure is a major challenge for scientists and researchers. The approach needs information from various sources and its use in the relevant field, which lead to a growing interest in knowledge discovery from large data. Data mining techniques provide effective solutions for this problem as it supports the automation of extracting significant data to obtain knowledge and trends, the elimination of manual tasks, easier data extraction directly from electronic sources, and transfer to secure electronic system of documentation, which will increase the agriculture productions from same limited resources. In a nutshell, the aim of this chapter is to gain insight into the applications of data mining techniques in smart farming, which direction to employ sustainable agriculture and identify the challenges to be addressed.

The mechanization of the process creates agriculture-based jobs for farmers, providing financial support and facilitating affordable agriculture equipment and machineries. Fruits markets are subject of opportunity and it is important to the suppliers to identify the quality of fruits based on the ripeness level of fruits before selling out in order to get higher level of profit. The proposed framework is an Android application in native language of the farmer to help the jobless farmers to find agriculture-based jobs suitable to their skill set and receive investments from various investors across the country. Further, it finds investment for the needy farmers and create suitable agricultural employment for jobless farmers so that there is an increase in the progress in the field of agriculture. It also facilitates the farmers with advanced equipment for performing various agricultural tasks, obtains the land on lease, and determines various stages of ripeness of fruit and provides the information about the government project and funding facilities.

The reduction of post-harvest losses and value addition of the horticultural corps has attained the higher priority of the current research works. Grading is the major phase in post-harvest handling. Presently grading is done on the basis of observation and through experience. Various drawbacks associated with such manual grading are subjectivity, tediousness, labor requirements, availability, inconsistency, etc. Such problems can be alleviated by incorporating automation in the process. Researchers round the clock are working towards the development of technology-driven solutions in order to grade/sort/classify various agricultural and horticultural produce. With the motto of helping the researchers in the field of grading and quality assessment of fruits and other horticulture products, the present work endeavors the following major contributions: (1) a precise and comprehensive review on technology-driven solutions for grading/sorting/classification of fruits, (2) major research gaps addressed by the researchers, and (3) research gaps to be addressed.

Enormous agricultural data collected using sensors for crop management decisions on spatial data with soil parameters like N, P, K, pH, and EC enhances crop growth for soil type. Spatial data play vital role in DSS, but inconsistent values leads to improper inferences. From EDA, few observations involve outliers that deviates crop management assessments. In spatial data context, outliers are the observations whose non-spatial attributes are distinct from other observations. Thus, treating an entire field as uniform area is trivial which influence the farmers to use expensive fertilizers. Iterative-R algorithm is applied for outlier detection to reduce the masking/swamping effects. Outlier-free data defines interpretable field patterns to satisfy statistical assumptions. For heterogeneous farms, the aim is to identify sub-fields and percentage of fertilizers. MZD achieved by interpolation technique predicts the unobserved values by comparing with its known neighbor-points. MZD suggests the farmers with better knowledge of soil fertility, field variability, and fertilizer applying rates.

Foreword

I am very much honored to give the foreword on the edited book entitled *Modern Techniques for Agricultural Disease Management and Crop Yield Prediction* which is being published under Advances in Environmental Engineering and Green Technologies (AEEGT) book series of IGI Global, USA. World's food supply management is an emerging challenge. Crop diseases are the major reasons of devastation of crops and decreasing crop production. Future impact of diseases on sustainable crop yields has invited debates amongst researchers. Agriculture researchers and scientists are continually making effort to develop new computational methods which enables efficient use of agricultural resources i.e. irrigation water, soil and pesticides. Complex interactions between a host, a pathogen, and their environment constitute dynamics of diseases. These dynamics leads to one of the largest risks facing the long-term sustainability of agriculture.

This book provides a window to valuable information resources on agricultural disease management and crop yield prediction and covers almost the necessary components which need to be known by researchers of the concerned field. Each chapter is one of its kind of and leaves the readers with curiosity to explore more about the discussed concepts and experiment the new methodologies of agricultural research. Out of ten total chapters, two chapters summarize emerging topics of computational methods used in agriculture field i.e. Deep Learning and Computer Vision. Two chapters are state-of-art literature review on related technologies i.e. Image processing techniques with Smart Agriculture features. Remaining chapters are few exemplary presentations of new inventions in Expert System Design, Automatic Data Acquisition and Spot Disease Identification System, Smart Agriculture Framework "Agro Guardian", Technology LED Solutions for fruit grading and outlier detection along with crop yield prediction.

It is my hope and expectation that this book will provide an effective learning experience and referenced resource for academicians, researchers, botanists, both UG and PG learners to know the current trends and their scope especially in the allied topics of Precision Agriculture that uses advances computational methods such as to Data Mining Machine Learning, AI, Computer Vision, Deep Learning and so on. Overall, the book is focusing on agricultural diseases and crop yield prediction, which aids the farmers to detect the agricultural diseases and to increase their crop yield.

I congratulate personally to the editors, authors who have played a vital role in bringing the edited book. Ultimately for any completion of good edited book there should be proper coordination and cooperation between the team of editors, authors and reviewers. As per my observation, I noticed the cordial coordination in the different teams.

I am guarantee that the book will act as useful source of information for researchers and academician who are eager to contribute their research findings in the allied areas/topics of precision agriculture.

Rajanish Kamalakar Kamat
Shivaji University, India

Rajanish Kamalakar Kamat *has published over 150 plus papers in International journals of repute and presented equal number of papers at National and International Conferences. One of his papers was awarded fourth position in the International Competition organized by the Association of Information Science and Technology (ASIST), USA. Amongst these research papers, many of them are on the latest themes such as Data Sciences, VLSI Design and ICT based pedagogy. He has published 12 books though reputed publishing house such as Springer UK. His books are regarded as best sellers and found their place in 170+ overseas institutes of higher learning from USA, UK, Germany, Singapore etc. Twelve students have been awarded Ph.D. under his guidance and 11 more are working for their Doctorate. He has been instrumental in resource mobilization under different funding schemes from UGC, DST, MHRD to the tune of Rs. 10 Crores.*

Preface

Agriculture fields, its crops and their contributions in human health development are the true measurement of human civilization process. History is full of examples when advancements in agricultural technology have significantly contributed in building nations. China is the finest example where agricultural developments provided strong foundation for rapid economic growth. Despite the traditional nature of agriculture, newly developed and emerging technologies such as 3D mapping, nitrogen fertilizer and genetic modification have pushed the agriculture industry towards revolutionary changes. Agriculture Disease management has become an integral component of the today's modern agriculture and this field gaining high attention of the world. Various studies suggest that world's population also increasing at a rapid speed. Agricultural scientists and researchers are getting compelled for exploring more options for increasing agricultural productions. Various agricultural diseases tend to decrease the potential production levels of a crop as per its time of incidence and intensity. The disease management interventions are required to be taken up at different phases of cultivation of a given crop.

Now a days, the farming technology is largely based on silicon chips and its latest contributions in agriculture field. Digital sensors monitoring system and weather observation systems are considered as few of the most powerful developments in farming industry which has come directly from creative and innovative methods placed around the farmland, inside tools, and in the farmer's hands. Internet of Things (IoT) devices, modern sensors, chips and monitoring equipment are assisting farmers to improve efficiency and increase their crop yields.

Development of a number of fungicides and other chemical based pesticides has witnessed in past century. High level overuse of chemical pesticides to minimize various insect pests and diseases over the years has ruined many naturally occurring effective antagonistic microorganisms. After that, the pathogens made a fight back through the emergence of pesticide resistant strains in the agricultural ecosystems. Later on, scientists recognized that due to heavy pressure of plant protection chemicals, the pathogens incline to change their genetic profiles. It motivated scientists to change the strategy from 'pest control' to 'pest management'.

Smart agriculture system has emerged in past few years and has gained interest of agriculture scientists. It relies more on the prevention of disease occurrences instead of managing them after they occur. It focuses on dealing with the roots of a disease problem rather than dealing with the symptoms of the diseases after their occurrence. Further, Smart Agriculture Systems advocates that there may be more than one strategy needed to tackle a single pathogen or a group of pathogens and a single set of interventions can target a variety of pathogens.

THE CHALLENGES

There are two critical issues which need to be addressed in order to implement effective and efficient agriculture disease management systems. One, obtaining a correct diagnosis method and second is implementation of the correct method in right manner. Another crucial factor which needs to be addressed is to understand different patterns of disease occurrences i.e. which diseases will occur and at which conditions and degree of severity. There are many factors which influences the disease development in variety of ways i.e. genetics of the pathogen populations, hybrid/variety genetics, age of the plant at the time of infection, weather (e.g., temperature, rain, wind, hail, etc.), environment (e.g., soil, climate) and single versus mixed infections. Huge variations in above mentioned factors cause difficulty in diagnosis of plant diseases at the early stages of disease on individual plants as well as at the early stages of an epidemic. However, for many diseases symptoms do become diagnostic at some stage of disease development and a reasonable level of confidence can be placed in diagnoses based on these symptoms.

Identification of disease is the most imperative while addressing the controlling measures. Many studies found that because of wrong identification of cause of disease, diseases are not successfully controlled or have caused recurring problems in future. Diagnosis in right manner is essential, especially in cases when many fungicides have a narrow spectrum of activity.

There are three contributors in development of disease: a susceptible host plant, the pathogen and ecological circumstances which favors the disease to get developed. These three components collectively referred to as the three sides of the "disease triangle." Our aim must be towards reducing one or more sides of the triangle which will further result into reducing the amount of disease.

Important principles of plant disease management include the use of resistant cultivars, sanitation, sound cultural practices and often fungicides. A holistic or integrated approach to plant disease control is the best approach and is highly encouraged.

There are two types of modeling methods used which are used by researchers for yield forecasting: statistical models and process models. Statistical models rely on defining the relationships from historical data and apply the new mathematical models onto patterns to predict future of yield. These models are generally personalized for specific crops and/or specific regions. Statistical models are very useful approaches when it comes to predict how crop yield will change. However these methods are not being directly able to determine the growth of plant. These methods can provide key insights into changes of crop yield in future without being dependent on information of the definite factor/parameter values of specific crops. Most fundamental crop yield models are process models. These models heavily rely on two inputs i.e. the physiological characteristics of plants which include data on photosynthesis per-unit leaf area and other environmental variables like properties of soil. Because of their accuracy to determine the productivity of crops and to measure how variable affects crop yield, these are widely popular among researchers. Few of such popular models include CERES-Maize model, GAEZ model and SALUS model, CROPGRO-soybean model. Although these are only a few examples, they each typify a major limitation of process based yield models: they require large parameter sets and input variables which are often unavailable and/or difficult to fit at large spatial scales for diverse crops.

SEARCHING FOR A SOLUTION

In past few years, we have witnessed many interesting published issues related to challenges of crop production and emerging techniques for related disease management. For example, McCalla, A. F. (2001) has highlighted the significance of Sustainable Natural Resource Management and increasing crop production using computational techniques. He advocated that knowledge-based agricultural intensification in conjunction with modern science and biological techniques will serve the requirements of food of increasing population of earth.

The combination of environmental data associated to digital images and geographical information are the primary motivators for the development of new techniques and algorithms which can help out in increasing crop production and managing agriculture diseases effectively. Emergence of machine learning and deep learning methods will enable researchers and scientists to understand hidden relationships, patterns and correlations of crop yields that were previously unknown. Other emerging areas of computer science such as statistics, databases, artificial intelligence, pattern recognition, machine learning and visualization are also added advantages for researchers in agriculture domain.

Data acquisition technology has become so much advanced that time interval for gathering data has been minimized at its best by using remote sensing images and time series analysis of data. Traditional agrometeorological methods have improved their accuracy by using remote sensing data which is collected from various sources i.e. rainfall. Principal component analysis, frequency distribution, geostatistics, cluster analysis, Fourier transform, non-parametric statistics are the major techniques used by researchers in Agriculture domain. However, there are many questions which are still to be answered.

The chapters in the book covers the topics from the fundamentals to the advanced concepts of precision agriculture. The authors have really strived hard to document a meaningful and complete chapters.

The book addresses various topics ranging from issues and challenges in Smart Farming for sustainable agriculture to Data Mining Techniques for Outlier detection, removal and Management Zone Delineation for Yield Prediction. The book can be considered as a ready reference and can be recommended as reference text book in the college/university libraries. The book can be considered as useful source of information for the researchers, academicians and practitioners, who are keenly interested in making some key findings, contributions especially in the allied areas of precision agriculture.

ORGANIZATION OF THE BOOK

The book is organized into 10 chapters. A brief description of each of the chapters follows:

Chapter 1 presents an analysis of issues and challenges in Smart Farming for sustainable agriculture. It advocates that sustainable agriculture improves the biodiversity of the land and it is very essential to ordinate with the increasing demand for the food, climate change and degradation of the ecosystem in future.

Chapter 2 is a state-of-art review of literature on the Image Processing Techniques which are aiding Smart Agriculture. Also this chapter presents various image processing means that could come in handy for ameliorating the present scenario and shows their implied extension in the smart farming world.

Chapter 3 proposes a modern Expert System Design for Diagnosis of Diseases for Paddy Crop. It highlights various reasons for India's low paddy crop yield and their related problems. The authors has aimed to address this low crop yield issue by designing an expert system which helps the farmers by identifying and predicting the diseases for paddy crop to enhance crop yield and to reduce the supply and demand gap.

Chapter 4 summarizes applications of Deep Learning and Computer Vision in Smart Agriculture. This chapter gives a detailed explanation of different Deep Learning methods and tries to provide a basic understanding as to how these techniques are impacting different applications in smart agriculture.

Chapter 5 discusses the recent techniques in image-based plant phenotyping along with their limitations. Also, this chapter proposes a new plant segmentation method based on Wavelet and leaf count methods based on Circular Hough Transform and Deep learning model which are experimented with Computer Vision Problems in Plant Phenotyping (CVPPP) benchmark datasets.

Chapter 6 advocates the Automatic Data Acquisition and Spot Disease Identification System in Plants Pathology Domain. It presents an Agricultural Intelligence System which uses k-mean, PCA and ICA algorithms to identify diseases in plants. The implementation proves that the proposed method shows promising result on the basis of Histogram of Gradient (HoG) features.

Chapter 7 analyses and compares the different applications of Data Mining Techniques in Smart Farming for Sustainable Agriculture. This chapter gives a detailed explanation of different Deep Learning methods and tries to provide a basic understanding as to how these techniques are impacting different applications in smart agriculture.

Chapter 8 presents the notion of "Agro Guardian" which is a Smart Agriculture Framework for Precision Farming. The proposed framework is an android application in native language of the farmer to help the jobless farmers to find agriculture based jobs suitable to their skill set. It also facilitate the farmers with advanced equipment's for performing various agricultural tasks, obtain the land on lease and to determine various stages of ripeness of fruit and provide the information about the government project and funding facilities.

Chapter 9 presents a study on Technology-LED Solutions for Fruit Grading to Address Post-Harvest Handling Issues of Horticultural Crops. With the motto of helping the researchers in the field of grading and quality assessment of fruits and other horticulture products, the present work endeavors the following major contributions: (1) A precise and comprehensive review on technology-driven solutions for grading/sorting/classification of fruits. (2) Major research gaps addressed by the researchers and (3) Research gaps to be addressed.

Chapter 10 presents an optimized Data Mining Techniques for Outlier detection, removal and Management Zone Delineation (MZD) for Yield Prediction. It uses Iterative-R algorithm for outlier detection to reduce the masking/swamping effects. MZD suggest the farmers with better knowledge of soil-fertility, field-variability and fertilizers applying rates.

Acknowledgment

The editors of the book wish to acknowledge sincerely to the entire authors who have contributed their wonderful chapters to our edited text book in the stipulated time. We also thanks again for the authors who have spent their precious time in fine tuning the chapter after receiving constructive comments from reviewers.

The editors will be thankful to everyone in the Editorial Advisory Board Team whose initial discussions helped us to bring out this good edited book.

None of the human beings are 100% perfect. Everywhere there is a scope of improvement. So, we thank entire reviewers' team for providing the constructive comments to the authors to improve the chapter with respect to quality, coherence and content presentation of chapters. Without reviewers support, this book would not have become a reality.

Few of the authors also served as referees; we highly appreciate their double task.

We believe in "Cooperation, Coordination, Commitment can make any project to be success". For the successful completion of the edited book, we the editors will acknowledge everyone who helped us directly and indirectly.

N. Pradeep
Bapuji Institute of Engineering and Technology, India

Sandeep Koutish
Guru Kashi University, India

C. R. Nirmala
Bapuji Institute of Engineering and Technology, India

Vishal Goyal
Punjabi University Patiala, India

Sonia Abdellatif
Seoul National University, South Korea

Chapter 1
Issues and Challenges in Smart Farming for Sustainable Agriculture

Immanuel Zion Ramdinthara
Pondicherry University, India

Shanthi Bala P.
Pondicherry University, India

ABSTRACT

Sustainable agriculture helps to promote farming practices and methods in order to sustain farmers and resources. It is economically viable, socially supportive, and economically sound. It assists to maintain soil quality, reduce soil erosion and degradation, and also save water resources. Sustainable agriculture improves the biodiversity of the land and thus leads to the healthy and natural environment. The sustainable agriculture is very essential to ordinate with the increasing demand for the food, climate change, and degradation of the ecosystem in future. It plays a major role for preserving natural resources, reducing greenhouse gas emissions, halting biodiversity loss, and caring for valued landscapes. Sustainable agriculture is applied to farming in order to preserve the nature without compromising the quality of the future generation basic needs and thus enable to make smartness in farming. The common practices included in smart farming for sustainable agriculture are crop rotations that mitigate weeds, disease, insect, and other pest problems.

INTRODUCTION

The word “sustainable” is the process of maintaining changes in the environment. Sustainable agriculture is a measure in which it should emphasize long-term support in producing food and beverages and also at the same time in an eco-friendly mannered (Srisruthi, Swarna, Ros, & Elizabeth, 2016). Sustainable agriculture helps to balance the needs for food with ecological preservation. It helps to promote farming

DOI: 10.4018/978-1-5225-9632-5.ch001

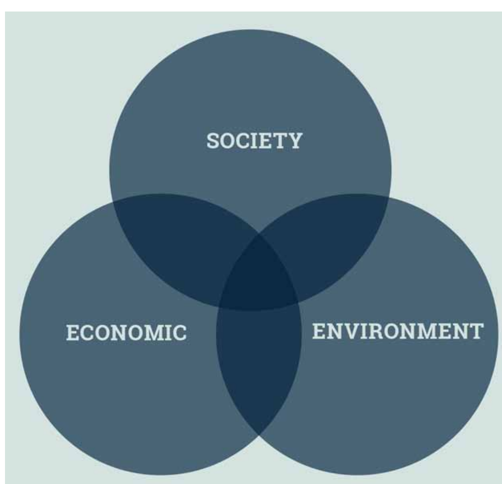
practices and methods in order to sustain farmers and resources. It is economically viable, socially supportive and economically sound. It assists to maintain soil quality, reduce soil erosion and degradation and also save water resources. Sustainable agriculture improves the biodiversity of the land and thus leads to a healthy and natural environment. Sustainable agriculture is important ordinate with the increasing demand for the food and control climate change and degradation of the ecosystem in the future. It plays a major role in preserving natural resources, reducing greenhouse gas emissions, halting biodiversity loss and caring valued landscapes.

Sustainable Agriculture

Sustainable agriculture is applied to farming in order to preserve nature without compromising the quality of the future generation basic needs and thus enable to make smartness in farming. The common practices included in smart farming for sustainable agriculture are crop rotations that mitigate weeds, disease, insect and other pest problems. Thus it leads a way to make the hazardless environment. Sustainable agriculture comprises of sustainability of farmers, the productivity of agricultural resources and environment-friendly. Recycling and harvesting of water is the milestone to form sustainable agriculture. Moreover, as the need for food rises every day, it has to be preserved in order to meet the need sufficiently.

The living organisms are dependent on the nature of biodiversity. This has been contaminated slowly by emitting wastes, degraded dead plants, use of fertilizers and pesticides, dilution of water, etc. Moreover, because of this desecration of an environment and emission of greenhouse gases actually affects the plants, animals as well as human beings. So, it is very important to sustain and make a better environment for plants and for human beings. Thus, smart farming is a primary key to meet better agricultural systems and make better sustainable agriculture for the future.

Figure 1. Three factors for sustainable agriculture



Smart Farming

The sustainability of agriculture can be carried up with a technique called smart farming. Smart farming is basically a concept of promoting precision agriculture and create eco-friendliness to increase the quantity and quality of the nourishment with modern sophisticated technology (O'Grady & O'Hare, 2017). Smart farming enables the farmers to monitor and control over the plants remotely and satisfy their necessary requirements of the plants and animals. Internet of Things (IoT) is a new technology that enables the devices to connect remotely in order to achieve smart farming (Patil & Kale, 2016). Smart farming delivers simplicity for the farmers to harvest and yield crops as automation of sensors and machines that replaces the early typical workforce of farming and it is much faster and reliable than ever before. The technologies transform the typical way of farming to automated devices which brings revolution in the history of agriculture.

Farming was completely dependent on labor force until it meets new technology in recent years. With the advancement in technologies, farmers could possibly make higher production and it has a great impact on the agricultural economy. It also helps in bridging the gap between the small and large-scale businessman. The technologies brought an ability to communicate with all over the world and explore what is going on on the other side of the world. This way methods and techniques can be shared by farmers across the globe.

Before the technologies come into existence, farmers and labors face a lot of problems in selling out the products that had been harvested as they hardly get the secondary buyers. Fortunately, even if there are buyers for the products, the farmers could not sell it off at a reasonable price. However, the emergence of technology had swiped away the nightmare that has held the farmers in misery. Agriculture has met a new technology which actually changed the typical way of farming and conventional techniques were transformed into a technique called the Internet of Things. This technology has drastically changed the way of farming and has much more potential of establishing better precision agriculture.

Internet of Things

The Internet of things is the interconnection of different devices over the internet through the cloud server. IoT is implemented and deployed in different platforms like Hospital, Traffic, Government Offices, vehicle, and agriculture, etc. (Asghari, Rahmani, & Javadi, 2019). There is a tremendous impact on agriculture and has assisted human labors and promotes simplicity (Khanna & Kaur, 2019). The technology has the ability to monitor the plants and animals and also can retrieve information remotely in the device like handheld and mobile phones. Moreover, the primary obstacles that the farmers faced to get a productive agriculture product are unpredictable weather, water scarcity, pests, and diseases. Unpredictable weather is apparently caused due to the pollution emitted by the human from industrial wastes and smoke from vehicles etc. Agriculture is completely dependant on the weather condition as it requires rainwater and light for photosynthesis. However, because of the pollution and wastes from industries and individuals, the weather of the season became uncertain and farmers faced problems because of unpredictable weather which affect the crops and falling of productivity.

Devices and sensors enable the farmers to predict weather and anticipate the amount of production. Secondly, Water scarcity is also the most common problem faced by farmers. This could also be more or less resulted by pollution and global warming. Water had been tremendously contaminated by industrial wastes and human daily wastes. Moreover, most of the inhabitants simply waste water and dump the waste in the sea. However, the people came to realize that the earth is in critical as summer became hotter and hotter and winter becomes extremely cold. This is a big sign of the consequences of human action. So, awareness had been given in educational institution and society to save rainwater and maintain waste properly. This is where technologies are highly important as they can be made in such a way it would assist human in any means. In many developed countries like USA, Israel and some of the European Countries, IoT plays a vital role in everyday living as even garbage trucks are automated to identify and differentiate degradable and nondegradable substances and wastes.

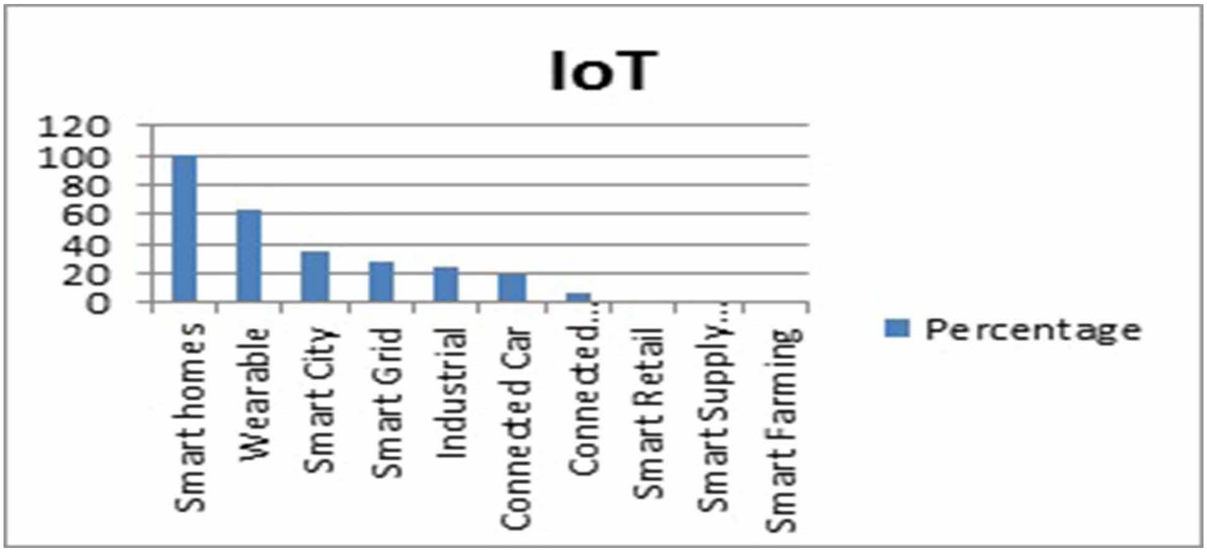
IoT has significantly played a very important role in the harvesting of water as it monitors the amount and controls the flow, it evaluated the amount of water required by the plants and supply them an adequate amount which saves a huge amount of water than ever before(Yong et al., 2018). Conventional water sprinkler was used in lawns and gardens which also saves a fair amount of water as compared to the manual methods of watering plants. Moreover, technologies become even better and smarter. Water sprinkler could not only supply water but also controls with different parameters prior to the humidity of the soil, grass, and plants. Sensors are connected to the cloud through the gateway and could be able to monitor the status and supply water exactly what the soil and plant need remotely through handheld devices and computers (Mekala & Viswanathan, 2017). Moreover, other means of harvesting water is drip irrigation. This has been widely used in many parts of the world. Drip irrigation is the method of dropping water directly to the root of the plants to an adequate amount (Joshi & Ali, 2017). This technology is used a wide and huge piping method which would supply many plants in large scale.

One of the problems that farmers encountered other than water scarcity is the disease and pests which destroy their crops and plants which keeps the farmers in devastation. It is difficult for the farmers to anticipate the invasions of the diseases as these are small and tiny substances which cannot be seen especially plants in a wide area of land. The farm production apparently deteriorated as the diseases had invaded many plants in the farms and the farmers could barely monitor every plant. It was problematic for the farmers to monitor and observe each and every plant manually whether it had suffered from disease or not. However, IoT technology fortunate farmers and lead them to a new milestone in the agricultural field. This technology has transformed the conventional way of farming to a new era of farming (Mittal & Singh, 2007).

Machine Learning

A machine Learning algorithm is basically a subset of Artificial Intelligence which helps to train and to learn automatically and improve the experience without any human consent and without explicitly programmed. This technology is thriving in the agriculture sector as it could perform a task automatically and also improved the technology of IoT. Machine Learning has become very powerful, accurate and efficient because it can make the best feasible solution for many situations. The term "Machine Learning" was coined by Arthur Samuel in the year 1959. It was used in several problems solving. Improvement takes place over time, it became a very powerful algorithm or system where the machine can learn and act accordingly and smartly. Machine Learning can be broken down in three subgroups called Supervised Learning, Unsupervised Learning and Reinforcement Learning. Supervised Learning is basically a machine learning algorithm that deals with labeled datasets. Whereas in unsupervised learning, the datasets are scattered, unknown, unclassified and it can be any group of different variables. In Reinforcement learning, the software has to take action in the environment and acquire the most feasible solution. It is like a trial and error approach.

Figure 2. User's mostly browsed IoT on the internet
(Source: www.i-scoop.eu)



The weather has a huge dependency for farming so it is necessary for the farmers to monitor and predict for the upcoming season which would ideally significant for the plants (A. Toreti, A. Maiorano, G. De Sanctis, H. Webber, A.C. Ruane, D. Fumagalli, A. Ceglar,S. Niemeyer, M. Zampieri, 2019). Fortunately, Machine Learning has the capability to predict the weather from previous records or datasets using regression techniques. This is quite powerful for prediction but unfortunately human has emitted pollution which makes the weather hard to predict as it is unbalanced.

Farmers are opportune with the ability to classify plants variations, diseases, and pest which is aid by the machine learning algorithm called classification. Nowadays, many species are classified using the sensor which helps the farmers to identify the plants and supply fertilizers and water accordingly. It also can classify diseases and pest within the leaves which helps the farmers to save the plants or trees at the early stages.

The incorporation of Machine Learning and the Internet of Things had made a revolutionary in the agriculture domain and thriven the technology to the next generation. Farmers are able to monitor and control the devices and sensors accurately and make a hike on agricultural production. However, there are some limitations of these two techniques which are basically in the analysis of data and records. These technologies are powerful for the smaller datasets and a small group of parameters, however, when the data becomes extremely large, it does not work quite well and often leads to redundancy and inaccurate of data manipulation. Data Analytics is the process of organizing and modeling a set of data in order to be able to manipulate these data accordingly. This technology has improved the dataset generated by the sensors and stored in a proper format. With this data, a problem can be solved and the decision could be made from the data stored.

RELATED WORK

Sustainable agriculture is a measure in which farming would be taken in a smarter manner and eco-friendly. Many research is carried out in the agricultural domain and it is drastically evolving as compared with the earlier conventional type of farming. A focus is on the Internet of Things and Machine Learning algorithm as these two technologies are the most powerful and most feasible algorithm for solving a problem. As agriculture is important for livelihood, it is necessary to take a challenge for the researcher, students, and enthusiasts in order to develop the farming system, irrigation system, and marketing system, etc.

Water is a primary need for the plant's growth. So, it is necessary to aware of the preservation of water with the new technologies like automatic water sprinkling systems, irrigation system and desalination of sea water. Some countries like Israel and the US have adopted and practices water preservation by implementing drip irrigation and desalination. Drip irrigation is basically a watering system used in agriculture for watering plants economically and also sufficiently. It saves water as it drips within a fraction of time and does not waste. It is a big challenge to improve the conventional drip irrigation technique and to perform it in a smarter way in which it can actually monitor the soil moisture and plants humidity (Pandithurai, Aishwarya, Aparna, & Kavitha, 2017). Over the year, water had been wasted due to improper water systems and preservations. In many countries, water scarcity is a major problem for the development of agriculture. Now, a sensor technology helps to monitor the pH of the soil, soil moisture and nitrogen content of the soil for the plants to consume the exact required amount of water and mineral.

Water sprinkler is another way of conserving water and watering plants economically for the adequate use of water. In recent decades, a conventional water sprinkler was used where it supplies water to the plants by sprinkling over the plants under the control of the operator. It is a semi-automatic in which the operator controls the switch where it could be turned on and off with operator acknowledgment.

IoT and Machine Learning are used in many agricultural platforms for smarter and better systems for upcoming experiences. So, these technologies can assist the farmers in monitoring the humidity of the soil, the pH level of the soil, and minerals, etc. The characteristics of all plants differ from one another, for instance, the amount of water required to grow and kinds of soil mineral and soil pH to have proper growth. So, in order to grow plants and yield crops faster, it is very important to implement a system where each plant gets its requirements. Sensors like Arduino, RasberryPi, and Zigbee are the most common devices used by many authors for monitoring the soil. These sensors are actually electronic components which can generate data from the sensor in the form of magnetic waves and current (Kalaivani, Allirani, & Priya, 2011). Data and information can be accessed and processed through the cloud server from anywhere on handheld devices for monitor and control. In this way, specific plant requirement can be measured and supply an adequate amount of minerals and herbicide.

During the late twentieth century, trackers and animals like horse, buffaloes, and cows are the main sources of energy for farming. Especially tractor and machines run by steam is apparently, afforded only by some of the countries from the American continent and Europeans. In other parts of the world like the African continent and South East Asian continent, workforce and animals are the primary sources of energy for farming. When the technological revolution in agriculture began, the American continent and Europeans are the countries to lead and it was growing exponentially. Since the 21st century, technology thrived rapidly and the Internet of Things was introduced which becomes a very important technology for many domains. Machine Learning has also improved soon after the introduction of IoT. These two technologies completely change the way of conventional farming to digital farming.

So, in this modern age of technology, diseases and pests could be easily detected using sensors and image processing which is an application of IoT and Machine Learning. Machines are trained to identify which plants are growing well and healthy. Image processing is used for visual identification in classification, detection of diseases. It is basically the process of comparing the images that have been captured by the visual camera or Aerial drones with several sample pictures (Dimitriadis & Goumopoulos, 2008).

Image processing is generally used for the plant disease detection that helps farmers to prevent the deterioration of plants and crops due to the pests and diseases (Gandhi, Nimbalkar, Yelamanchili, & Ponkshe, 2018). Plant disease is one of the most destructive measures in the agricultural sector which often brought the farmer in misery. An image-based classification system for plants diseases (Athani, Tejeshwar, Patil, Patil, & Kulkarni, 2017). Datasets are taken manually from the input device and to augment the input dataset, Generative Adversarial Network is used and further classified using Convolutional Network (CNN). In agriculture, water preservation is one of the vital measures that must be taken in order to sustain the life of plants and trees. They have introduced a system that measures the soil moisture and humidity using sensor and IoT technology. The main objective is to utilize the water for irrigation. The plant would be watered adequately without wasting any of the water in order to meet the needs of plants as well as to save the water. The Arduino sensor is deployed in the soil and fetches information in the form of data. Soil moisture and soil pH level is measured and processed further using the Neural network algorithm.

Data Analytics is also a significant topic in modern technology as data is generated at every moment. Every enterprises, company, and machine significantly generated information or data to process and records all the transaction. Data Analytics is a science of analyzing a large group of unlabeled data and manage the trends, groups, and types to form a uniform database. Data Analytics and Big data are already implemented in many domains like offices, hospitals, and airport, etc (Lim, Kim, & Maglio, 2018). Big Data is a new approach used for managing and manipulating a massive group of raw data and store for future reference and for a better future experience. It could collect information from a different platform and organize them for comparison and survey (Wolfert, Ge, Verdouw, & Bogaardt, 2017).

TECHNOLOGIES FOR SUSTAINABLE AGRICULTURE

Technologies for Smart Farming

The labor force is the primary source of power and energy engaged in agriculture in the late 20th century although some conventional machines and animals are still in use. The farmers were often frustrated and devastated because of the unpredictable situation like weather change, water shortage, pests, diseases and calamities, etc. Season of the year was the only thing that could be anticipated by the farmers but, now even time of the season gradually changes because of global warming and pollution.

However, the 21st century brought the technological revolution which gradually changes the way of farming. The animals are used for farming like horses and buffaloes were replaced with tractor and machines. In the early days, machines were operated physically by farmers or operators which is through tremendous ease as compared to the labor force farming. The technological revolution brought tremendous changes to farming and increase crop productivity. The evolution had brought machines like a tractor, truck, combine harvester, etc which is called mechanization. There are different types of technologies which together makes farming smart. They are:

- Sensing technology
- Software application
- Information and Communication technology
- Positioning technology
- Hardware and Software system that enabled IoT-based
- Data Analytics

Sensing Technology

Sensing technology is basically a device which has the ability to measure certain properties of some components or variables. Sensing technology changes the typical ways of farming in many ways such as monitoring the plants, monitoring soil moisture, minerals and pH level. Weather forecasting became much accurate. Weather conditions are also one of the primary measures to be monitored as particular crops required to be grown inadequate temperature level. There are several brands for a sensor such as Flir, UNO, Raspberry Pi and Arduino, etc.

Software Application

A software application is a group of programs that can perform specific task and functions. It is the interpreter between the users and the hardware. Many software developers and communities develop several software applications which promotes simplicity in operation of hardware devices and machines to perform tasks without much of human assistance. The software application provides a privilege to the users to touch and control over the devices and machines and helps to perform a specific task. In farming, Software Application is the module which controls and operates over all the sensor and farming devices through handheld devices and computers.

Information and Communication Technology

Information and communication technology refers to a technology that provides access to data and information through telecommunication and internet interconnection. This technology allows the user to exchange data and information through network interconnection from device to another. There are different types of interconnection and communication such as Wireless Sensor Network (WSN), Radio Frequency Identification (RFID), and Zigbee network, etc. These types of communication are used in many domains such as Airport, Traffic, hospitals and even in agriculture. Wireless Sensor Network is a type of communication that has a standard of 802.15.4 that probably covers up to 100 meters depending on the settings and devices used. This is widely used for monitoring soil moisture, soil pH level, soil minerals, disease and pests and even weather prediction which is control through the cloud server (Abhiram Singh & T. P. Sharma, 2014).

A Radio Frequency Identification (RFID) technology refer to a wireless system that allows the device to read information from a certain distance without any physical contact. It provides a method to transmit and receive data from one point to another. This is widely used as a barcode reader and ID scanner, etc. RFID is also used for identifying animals in the livestock and for labeling the animals (Ruiz-Garcia & Lunadei, 2011).

Zigbee is of 802.15.4 standard which is a high-level communication protocol and it is used to create a small personal area network with low power and low bandwidth. This communication is generally used in a small area. It is suitable for farming within a specific area which connects the sensors and the controller.

Positioning Technology

Positioning technology is a technology for determining a position and orientation of an object or a thing. It is widely used today for determining the location on a map. Global Positioning System (GPS) and Global Information System are the two prominent systems which are used by many companies like Google Map for locating positions. In smart farming, positioning system plays a vital role in which the farmers enables to locate particular crops over the wide field when required which makes it faster. It also helps in identifying areas which are suitable for cultivation, etc.

Hardware and Software System That Enables IoT-Based

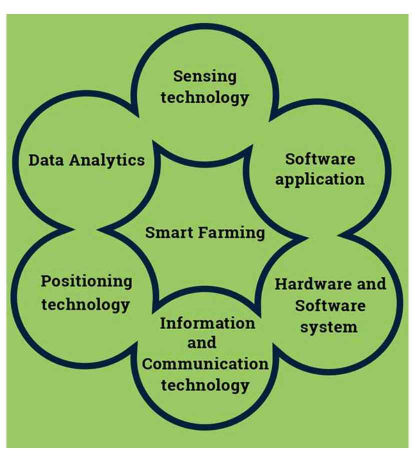
In the world of computing, hardware and software always come together which connects and communicate to perform a task. Hardware comprises a physical electronic substance like Devices and sensors. Software is the group of commands or program which makes the hardware works such as Arduino software, Machine Learning, and ThinkSpeak, etc.

Data Analytics

Data Analytics is the process of organizing a group of raw data into uniform sets of information which can be referred and manipulate for a better experience. In farming, as the farmer has to record information about different types of crops and their properties, data analytics can be helpful in manipulating the data as per required.

The figure above shows the important six factors that make farming smart.

Figure 3. Different factors making farming smart



With the growth of population globally, the need for food has also arisen respectively, so, it is necessary to aware of food security to meet the need. Moreover, due to the pollution and wastes, the form of season gradually changes and weather are improper as compare with the early days. The weather became unpredictable and farmers the most sufferers as plants required proper light and water for their growth. That is why smart farming is important to control all the problems much accurate.

Smart farming is the only feasible way to sustain agriculture and produce quality products at a higher quantity. It is highly efficient as compared with conventional techniques. IoT technology that provides uses sensors like sound, light, temperature, humidity and soil moisture which can be used as a measure to monitor the crop's and plant's growth and identify the requirements of fertilizers and minerals of the plants (Biradar & Shabadi, 2017).

Parameters for Smart Farming

The different parameters that are used in smart farming are as follows:

- Disease detection
- Smart water sprinkler
- Soil moisture
- Soil pH
- Soil minerals
- Soil temperature
- Water Sprinkler
- Drip Irrigation

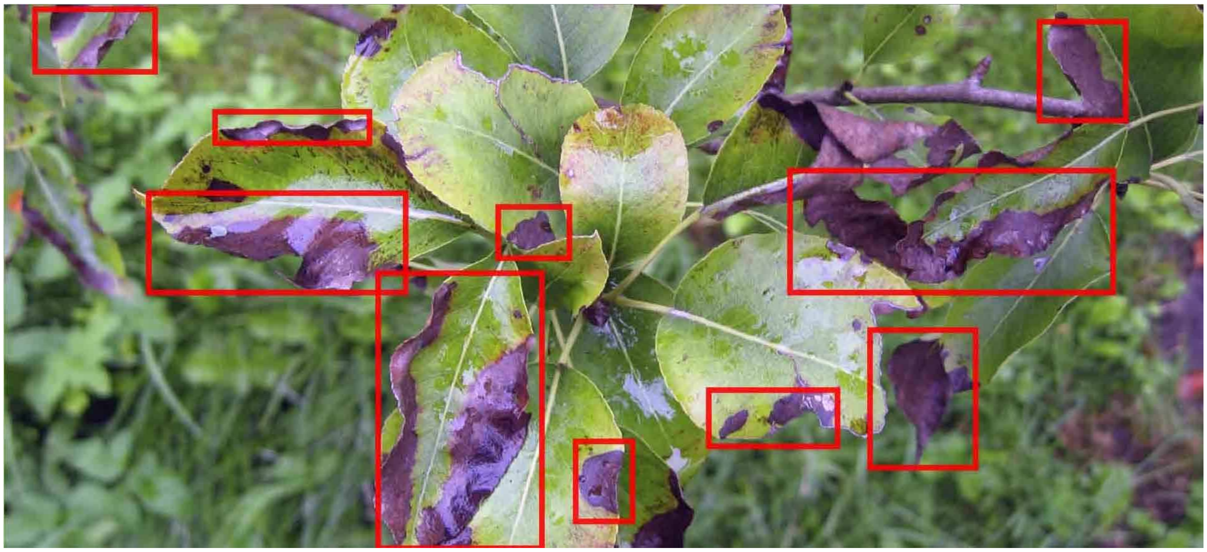
Basically, farming is improved since recent years with the use of sophisticated technologies which for faster work and profitable. Some of the technologies are

Disease Detection

During a period of harvesting and cultivation, farmers are often disturbed by pests and diseases. Plants disease are very common among different varieties of plants and it is usually caused by fungus, bacteria, and virus. The occurrence of disease in plants may differ from time to time. There is the season when diseases are likely to spread and attack plants which leads to decrease crop production rates. The sensor technology and Machine Learning help in monitoring the plant's health and detection of diseases at the early stages when plants leaves can be cured easily. Here, sensors like Visual sensors and camera are highly used for disease detection as it provides a visual picture for any detection which is much accurate and reliable. The camera is installed in such a way that it would scan the leaves real-time further processed with Machine Learning algorithm which would classify the pest and the disease found on the leaves (Singh, Varsha, & Misra, 2015).

These diseases are also detected with various devices like an electromagnetic sensor, Optical sensor, Mechanical sensor, electrochemical sensor, airflow sensor, and an acoustic sensor.

Figure 4. Disease detection with image processing technique
(Source: bitrefine.group)



Smart Water Sprinkler

The advancement in technology has also brought a technique for the preservation of water. In the earlier system of a sprinkling of water, the water sprinkler was controlled by the operators or farmers towards their knowledge and experiences to utilize and save water. However, in the present technology, the soil is being sensed initially using sensors which check the humidity and moisture of the soil and pass input parameter for the further process to control the water sprinkler depending on the inputs requirements. This technique is much more sophisticated and more efficient as compared with recent technology.

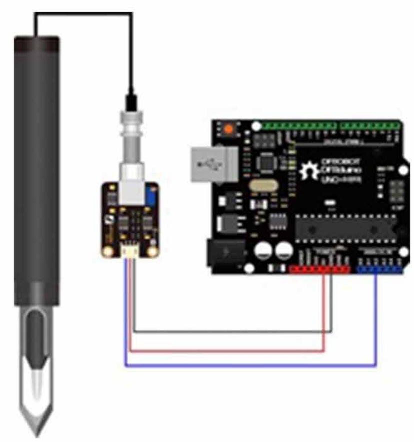
Weather Forecasting

Crop cultivation completely dependent on the weather condition as crops certain ranges of temperature where it can live and grow. So, it is crucial to control the farmers in order to increase the food production rate and maintain high food security. As human is emitting pollution to the environment, it also has a great effect on climate weather change. This has made farmers in misery as they are not able to anticipate for the upcoming weather. However, with the help of technology, it is possible to predict for the upcoming weather for the better farming experience. Sensors for weather forecasting are used widely in agricultural systems.

Soil pH Level

Soil pH is the measure of acidity and alkalinity in the soil. It consists of level 1 to 14 where 7 is the neutral level for the plants to adapt to live and grow. However, some plant has the ability to adapt and even thrived beyond the neutral level. So, it is very important to know the kind of soil so that the farmers may

Figure 5. UNO soil pH sensor



seed the particular plants and trees in order to thrive. IoT technology has changed the conventional way of measuring soil pH with chemicals. Sensors technology like Arduino are widely used for monitoring the soil pH with device probe dipped in the soil which receives a signal in the form of Radiofrequency.

The sensor which looks like a pen is dipped into the soil. It is functioning like potentiometer and signal or data information would be sent to the board for further processes.

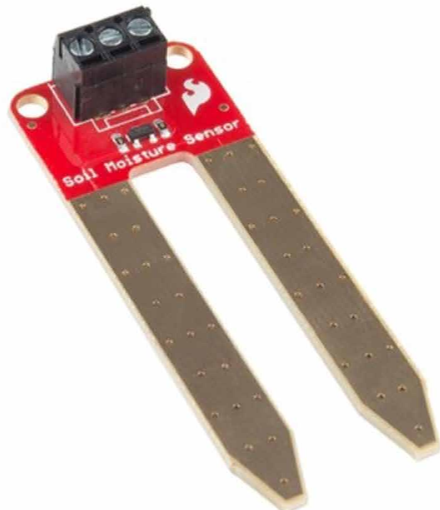
Soil Moisture

The number of water requirements for every plant could probably differ from plants to plants in order to grow. So, in order to supply a sufficient amount of water a probe sensor is used which measures the water content of the soil (Pandithurai et al., 2017). Arduino, Zigbee, and raspberry are the common sensors that are mostly used today. The soil moisture sensor has two probes which allow current to flow.

Soil Temperature

Soil temperature is the measurement of the warmth of the soil. Plants have an ideal temperature range to grow which is of 65-75 F (18-25C). However, there are several crops that have different ranges of temperature in which it can survive. In the early days, measuring the temperature of soil not possible so, it could be proof by monitoring the growth of plants. As the technologies emerged, however, the sensor technology enables the farmers to anticipate the conditions of the soil.

Figure 6. Soil moisture sensor



Soil Minerals

Soil contains a number of minerals, nutrients which are essential for the plants to grow like phosphorus denoted as (P), nitrogen denoted as (N) and potassium denoted as (K). These minerals help the plants to grow as it is the primary nutrients for the plants. On the hand, there is an excessive amount of these minerals present which actually lead to contaminate the groundwater. So, it has to be neutralized just enough for the plants and soil. Excessive use of fertilizers and herbicides is harmful to the crops and plants as well as for the consumers. So, it is important to avoid these chemicals and promote using of technologies for monitoring of soil mineral for better growth of the plants.

Figure 7. Mineral sensor

Drip Irrigation

Drip Irrigation is an irrigation system that provides the ability to save water by dripping waste slowly to the root of the plants which can be buried inside or on top of the soil surface. It has the potential to save water because the water is dropping within a certain period of time. It also helps in minimizing the evaporation of water. Drip irrigation was first introduced by Simcha Blass and his son Yeshayahu in 1959 in Israel which reduces the water consumption and increase crop yields and productivity. Drip Irrigation is the reason behind the success in the agriculture sector. Israel is a desert where the country has only 20% of arable land. Apparently, it is because of the drip irrigation and technology that lifted the country up to this level and became on the most advanced country in agriculture. Drip irrigation has a great impact on farming and it can possibly increase crop yields and productivity (Kavianand, Nivas, Kiruthika, & Lalitha, 2016). The figure shows how drip irrigation works. The water source provides water to the pipe where the there are small knob or outlet valve where it drips the water.

IoT AND MACHINE LEARNING FOR PRECISION AGRICULTURE

Significance of Incorporating IoT and Machine Learning

The Wireless sensor network technology has thrived in the technology world and it transforms the conventional ways of the manual system to an automatic smart system (Deepika & Rajapirian, 2016). Nowadays, station, office, home, airport, farming, and traffic interconnect with the help of IoT technology. This is one of the fastest growing technologies as it had connected 15.41 billion devices in 2015 and estimated that 75.44 billion IoT devices would be connected by 2025 (Source: www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/). Internet of Things provides a common platform for all the devices where information and data can be shared and also provide a common language for all the

Figure 8. Drip irrigation

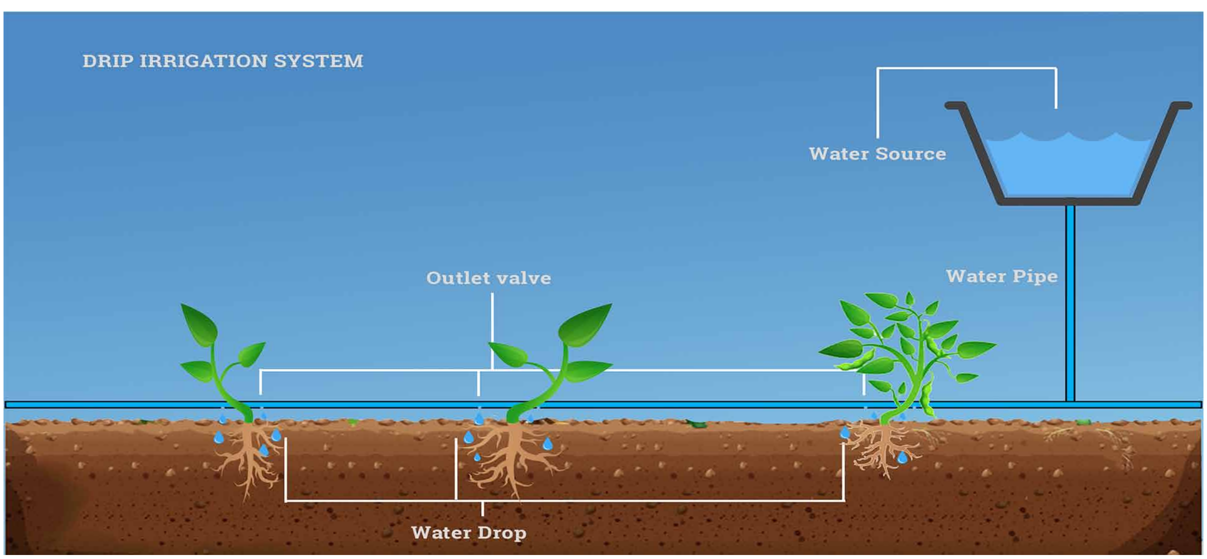
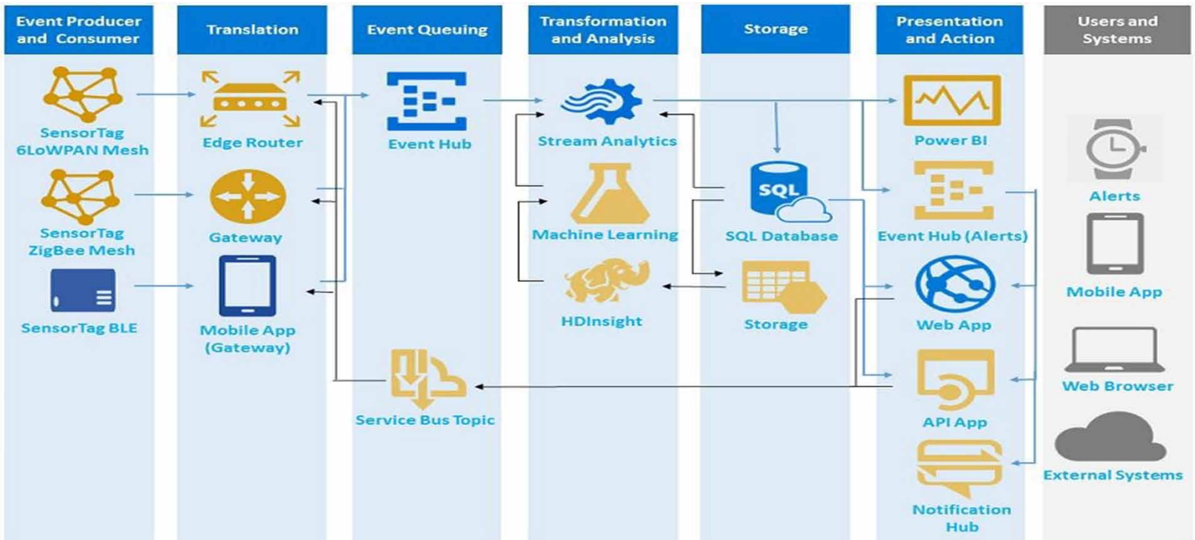


Figure 9. IoT and machine learning working mechanism
(Source: https://twitter.com/kirkdborne/status/762654729939869696)



devices for communication. Moreover, IoT provides accessibility from any platform and devices to all the information gathered and stored in a database. It provides authority to the users to retrieve the data as per the requirement for better experiences and it enables the users to monitor and control from any handheld device and computers. For instance, the fitness band is a new IoT technology that observes human gesture, behavior, and heartbeats which is actually a health monitoring device.

With the revolutionary in the system of farming, the Internet of Things (IoT) has been introduced which is basically the interconnection of all devices and sensors over the internet. This technology has drastically changed the life of a farmer as they can monitor and control over the farms and plants on the tips of their hand using handheld devices and laptops etc. This technology had made farming much faster and accurate and which ultimately lead to high productivity.

In this world of technology, IoT has emerges to wide ranges in almost every domain such as hospitals, traffic, vehicle, machines and even household which make technology smart. Technically, sensors are installed and are all connected all over the world through internet connectivity and data and information could be interchange and exchange from one domain to the other.

Machine Learning is an algorithm that teaches systems to work progressively for a better experience and make decisions from experience without explicitly programmed. It literally means training a computer to solve a certain level of real-time problems encountered in human life and without a human assistant. Machine Learning is a powerful system which is widely used in many domains such as a vehicle, agriculture, traffic, market and homes, etc. For instance, a self-driving car is one of the most prominent automated device built with a Machine learning algorithm. So, IoT and Machine Learning technologies have a high potential to create a revolutionary in the technology world. It is the combination of all the expert machines and intelligent system connecting together that enables to communicate and exchange information for a better experience.

Applications

Precision Farming

Precision farming is the measures that have taken to improve the system of farming with sophisticated technologies that provide efficiency and accuracy. Farming has been enriched with the technologies like sensors, automated machines, robots, autonomous vehicles and control systems. These technologies are used in different parameter and in different geographical places. These technologies could share data and information with high-speed internet connectivity. The Internet plays a vital role in the field of agriculture as it has the capability of controlling and bringing all devices together at one platform and share information. This connectivity has made farming effortless but also accurate and it has promoted control remotely.

In order to make precision farming, farming has to be smart, intelligent and also efficient and accurate at the same time.

Agricultural Drones

The drone has become a very important device in farming. It has been used for different parameters such as irrigation, disease detection, soil analysis, health assessment and so on. It has a very high time complexity and simplicity which can save a lot of time. With the drone, farmers could be able to monitor the crop, its health, yield prediction, plant height and amount of water requirement by placing camera which captures the image.

Livestock Monitoring

The animals in the livestock are large in numbers and it is very difficult for the farmers to monitor animal's health individually, behavior and diseases spread among the animals. Here, with the help of camera or sensor animal can be monitored real-time and retrieve images and data of the animal behavior and gesture anytime needed where it is very easy to identify the condition of animals. There are often spread of contagious diseases like H1N1, swine flu, fowlpox and bird flu which are easily spread amongst one another. So, monitoring of these animals would highly help to avoid the spreading of diseases amongst the animals by isolating the one with the disease.

Smart Greenhouse

The greenhouse is an enhancement of the system of farming in order to yield crops better and more productive. The main idea of implementing greenhouse is to maintain climate, weather, and humidity for the proper growth of the plants. In the greenhouse, plants and crops are growing rigorously as they are under proper environment. However, the conventional greenhouse has a limitation on automation where works are manually operated. To enhance the technology, smart greenhouse came into the picture with the help of IoT technology where it monitors the humidity, temperature, and climate specifically. IoT provides farmers with the ability to monitor and control over the internet through Wifi connections.

ISSUES AND PROBLEMS FOR IMPLEMENTING SMART FARMING

Technology has elevated the system of farming and has provided efficiency, accuracy and time complexity. Smart farming delivers an increase in productivity and yield crops. However, there are problems in adopting technologies in smart farming, these are:

High Cost of Technology

Recent technologies such as the Internet of Things and Machine Learning etc could minimize the workforce as it performs task really fast and accurate on the devices and machines etc, so, it is anticipated that the machine would probably replace the farmers in the near future. However, this is not really the case at some point as in some countries in the African continent and South East Asian continent because many countries had undergone through poverty where the workforce was the main source of energy in the agricultural fields. Therefore, deployment and implementation of devices and technologies are still not in the process. Implementation of devices practically on the field would probably require couples of sensors which would actually cost a huge amount of money. So, there are fewer chances of mechanizing the system of farming for the farmers while there are times when they only get just their daily bread. It would have been difficult for them to afford these kinds of devices while they are still having difficulties in implementing conventional tools for yielding of crops, productivity, and exports.

Unreachability of Rural Areas

Agriculture and farming take place in the countryside and isolated area where spaces are available for farming. It is because farming is more effective in a place where there the land is much arable than the land contaminated by the wastes and chemicals emits by the human. However, the implementation of technologies in these areas could be problematic as electricity and network coverage area are limited to the remote and rural area. So, people living in these areas often had a problem with electricity and power.

Immediate implementation of technology in farming would be difficult as farming needs huge acres of land and there are places in rural areas where internet connection and electricity has not reached in.

Ignorance by the Authorities

Farming typically takes place in the remote areas and mountainous regions where crops can easily adapt and soil compatibility is higher. Geographically, it has a high potential for yielding crops on a large scale with high productivity. However, the people who are actually farmers living in this area are poor. They are generally ignored by the authorities and hardly get financial support. The authorities support is probably the only way for the farmers to sustain agriculture. Farmers in these places do not have the capability of affording sophisticated farming devices and machines. Despite being mechanizing the system of farming, they could not even buy fertilizers and pesticides. Farmers from South Asian and South East Asian generally face these problems (Babar Shahbaz, Tanvir Ali1, Izhar A. Khan and Munir Ahmad, 2010).

Lack of Financial Resources

A group of financial supporters like governments and private banks and private money lenders could not give loans to the farmers because of several loans not being paid. This is because in some cases, the farmer could not get expected yield productions because of several calamities like droughts, Storms, unexpected increase in temperature and flood. Moreover, pests and diseases are also destroying the crops and it has been a nightmare for the farmers. So, many farmers eventually ended up with nothing in their hands. So, in many countries from the African continent and South East Asian continent farmers are poor and frustrated and even committed suicide.

Lack of Knowledge

Farmers in developing countries are mostly uneducated and unskilled because neither they are an urge to acquire knowledge of new technologies nor the authorities give awareness of the importance of new technologies (J. M. Kimiti, D. W. Odee & B. Vanlauwe, 2009). So, this is the main factor why farmers prefer the conventional type of farming over smart farming as it needs less money to spend (Abdul Rasheed Khan, M.K. Dubey, P.K. Bisen and K.K. Saxena, 2007).

A Development Project in Nigeria called FADAMA is an objective which will increase the incomes for farmers of rural areas and develop sustainable agriculture. It actually helps the farmers by financing them and makes them aware of the knowledge in anyways in order to develop and sustain a better way of farming, irrigation and increase food security in the country.

FUTURE RESEARCH DIRECTIONS

Internet of things is a technology that thrived on a very large scale and it enlarged its boundary in many domains, especially in the agricultural sector it has promoted simplicity for the farmers and increases yield productions. The estimation states that by 2020, over 24 billion devices would be connected with IoT. So, as the connectivity increases, security should go along the pace of new technology to protect systems from redundancy and several errors. This is also a big challenge for researchers and students to be aware of better security systems. Moreover, on the other side of the technology world, machine learning has grown so fast and is adopted in many. There are lots of researchers and enthusiasts people in many parts of the country seeking to enhance and make computer smart like a human. Incorporating of machine learning algorithms and IoT has the huge potential and feasible on the agriculture sector for making precision agriculture. So, a researcher, a student and enthusiasts must aware of these technologies and bring these technologies to its best for making a better smart farming experience.

CONCLUSION

In this chapter, different types of techniques which were used for farming and the modern sophisticated technologies are discussed and how it has an impact on farming and the overview of all the technologies. Apparently, the population growth rate is increasing exponentially which led the demand for food has risen respectively, so it is necessary for everyone to aware of the food security in order to sustain precision agriculture. Pollution and waste contaminate seawater, rivers, lakes and even the soil which actually affects human livelihood. The lakes and rivers water is the main source of water for consumption and for irrigation. Because of this contamination of these water, diseases like bacteria and viruses are often spread in human and plants, which is very dangerous for health. So, it is very important and crucial to emphasize on implementation of technologies irrigation, agriculture and drinking water in order to sustain mental well being. As agriculture is necessary for human livelihood, it is our duty to seek a better way of farming with modern technology to sustain Precision Agriculture.

ACKNOWLEDGMENT

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

Asghari, P., Rahmani, A. M., & Javadi, H. H. S. (2019). Internet of Things applications: A systematic review. *Computer Networks*, *148*, 241–261. doi:10.1016/j.comnet.2018.12.008

Athani, S., Tejeshwar, C. H., Patil, M. M., Patil, P., & Kulkarni, R. (2017). Soil moisture monitoring using IoT enabled Arduino sensors with neural networks for improving soil management for farmers and predict seasonal rainfall for planning future harvest in North Karnataka — India. In *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)* (pp. 43–48). Palladam, Tamilnadu, India: IEEE. 10.1109/I-SMAC.2017.8058385

Biradar, H. B., & Shabadi, L. (2017). Review on IOT based multidisciplinary models for smart farming. In *2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)* (pp. 1923–1926). Bangalore: IEEE. 10.1109/RTEICT.2017.8256932

Deepika, G., & Rajapirian, P. (2016). Wireless sensor network in precision agriculture: A survey. In *2016 International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS)* (pp. 1–4). Pudukkottai, India: IEEE. 10.1109/ICETETS.2016.7603070

Dimitriadis, S., & Goumopoulos, C. (2008). Applying Machine Learning to Extract New Knowledge in Precision Agriculture Applications. In *2008 Panhellenic Conference on Informatics* (pp. 100–104). Samos, Greece: IEEE. 10.1109/PCI.2008.30

Gandhi, R., Nimbalkar, S., Yelamanchili, N., & Ponkshe, S. (2018). Plant disease detection using CNNs and GANs as an augmentative approach. In *2018 IEEE International Conference on Innovative Research and Development (ICIRD)* (pp. 1–5). Bangkok: IEEE. 10.1109/ICIRD.2018.8376321

Joshi, A., & Ali, L. (2017). A detailed survey on auto irrigation system. In *2017 Conference on Emerging Devices and Smart Systems (ICEDSS)* (pp. 90–95). Mallasamudram, Tiruchengode, India: IEEE. 10.1109/ICEDSS.2017.8073665

Kalaivani, T., Allirani, A., & Priya, P. (2011). A survey on Zigbee based wireless sensor networks in agriculture. In *3rd International Conference on Trendz in Information Sciences & Computing (TISC2011)* (pp. 85–89). Chennai, India: IEEE. 10.1109/TISC.2011.6169090

Kavianand, G., Nivas, V. M., Kiruthika, R., & Lalitha, S. (2016). *Smart drip irrigation system for sustainable agriculture. In 2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR)* (pp. 19–22). Chennai, India: IEEE; doi:10.1109/TIAR.2016.7801206

Khan, A. R., Dubey, M. K., Bisen, P. K., & Saxena, K. K. (2007). Constraints faced by farmers of Narsing Kheda village of Sihore district. *Young (up to 30 yrs.), 8*, 16.

Khanna, A., & Kaur, S. (2019). Evolution of Internet of Things (IoT) and its significant impact in the field of Precision Agriculture. *Computers and Electronics in Agriculture*, *157*, 218–231. doi:10.1016/j.compag.2018.12.039

Kimiti, J. M., Odee, D. W., & Vanlauwe, B. (2009). *Area under grain legumes cultivation and problems faced by smallholder farmers in legume production in the semi-arid eastern Kenya*. Academic Press.

Lim, C., Kim, K.-J., & Maglio, P. P. (2018). Smart cities with big data: Reference models, challenges, and considerations. *Cities (London, England)*, *82*, 86–99. doi:10.1016/j.cities.2018.04.011

Mekala, M. S., & Viswanathan, P. (2017). A Survey: Smart agriculture IoT with cloud computing. In *2017 International conference on Microelectronic Devices, Circuits and Systems (ICMDCS)* (pp. 1–7). Vellore: IEEE. 10.1109/ICMDCS.2017.8211551

Mittal, A., & Singh, A. (2007). Microcontroller based pest management system. In *Second International Conference on Systems (ICONS'07)* (pp. 43–43). Martinique, France: IEEE. 10.1109/ICONS.2007.35

O'Grady, M. J., & O'Hare, G. M. P. (2017). Modelling the smart farm. *Information Processing in Agriculture*, *4*(3), 179–187. doi:10.1016/j.inpa.2017.05.001

Pandithurai, O., Aishwarya, S., Aparna, B., & Kavitha, K. (2017). Agro-tech: A digital model for monitoring soil and crops using internet of things (IOT). In *2017 Third International Conference on Science Technology Engineering & Management (ICONSTEM)* (pp. 342–346). Chennai, India: IEEE. 10.1109/ICONSTEM.2017.8261306

Patil, K. A., & Kale, N. R. (2016). A model for smart agriculture using IoT. In *2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC)* (pp. 543–545). Jalgaon, India: IEEE. 10.1109/ICGTSPICC.2016.7955360

Shahbaz, B., Ali, T., Khan, I. A., & Ahmad, M. (2010). An analysis of the problems faced by farmers in the mountains of Northwest Pakistan: Challenges for agri. extension. *Pakistan Journal of Agricultural Sciences*, *47*(4), 417–420.

Singh, A., & Sharma, T. P. (2014, July). A survey on area coverage in wireless sensor networks. In *2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)* (pp. 829-836). IEEE. 10.1109/ICCICCT.2014.6993073

Singh, V., Varsha, & Misra, A. K. (2015). Detection of unhealthy region of plant leaves using image processing and genetic algorithm. In *2015 International Conference on Advances in Computer Engineering and Applications* (pp. 1028–1032). Ghaziabad, India: IEEE. 10.1109/ICACEA.2015.7164858

Srisruthi, S., Swarna, N., Ros, G. M. S., & Elizabeth, E. (2016). Sustainable agriculture using eco-friendly and energy efficient sensor technology. In *2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)* (pp. 1442–1446). Bangalore, India: IEEE. 10.1109/RTEICT.2016.7808070

Toreti, A., Maiorano, A., De Sanctis, G., Webber, H., Ruane, A. C., Fumagalli, D., ... Zampieri, M. (2019). Using reanalysis in crop monitoring and forecasting systems. *Agricultural Systems*, *168*, 144–153. doi:10.1016/j.agsy.2018.07.001 PMID:30774182

Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M.-J. (2017). Big Data in Smart Farming – A review. *Agricultural Systems*, *153*, 69–80. doi:10.1016/j.agsy.2017.01.023

Yong, W., Shuaishuai, L., Li, L., Minzan, L., Ming, L., Arvanitis, K. G., ... Sigrimis, N. (2018). Smart Sensors from Ground to Cloud and Web Intelligence. *IFAC-PapersOnLine*, *51*(17), 31–38. doi:10.1016/j.ifacol.2018.08.057

ADDITIONAL READING

Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture*, *145*, 311–318. doi:10.1016/j.compag.2018.01.009

Goap et al. - 2018 - An IoT based smart irrigation management system us.pdf. (n.d.).

Ip, R. H. L., Ang, L.-M., Seng, K. P., Broster, J. C., & Pratley, J. E. (2018). Big data and machine learning for crop protection. *Computers and Electronics in Agriculture*, *151*, 376–383. doi:10.1016/j.compag.2018.06.008

Ma, C., Zhang, H. H., & Wang, X. (2014). Machine learning for Big Data analytics in plants. *Trends in Plant Science*, *19*(12), 798–808. doi:10.1016/j.tplants.2014.08.004 PMID:25223304

Mohamad Noor, M., & Hassan, W. H. (2019). Current research on Internet of Things (IoT) security: A survey. *Computer Networks*, *148*, 283–294. doi:10.1016/j.comnet.2018.11.025

Mostafa, H., El-Nady, R., Awad, M., & El-Ansary, M. (2018). Drip irrigation management for wheat under clay soil in arid conditions. *Ecological Engineering*, *121*, 35–43. doi:10.1016/j.ecoleng.2017.09.003

Rehman, T. U., Mahmud, M. S., Chang, Y. K., Jin, J., & Shin, J. (2019). Current and future applications of statistical machine learning algorithms for agricultural machine vision systems. *Computers and Electronics in Agriculture*, *156*, 585–605. doi:10.1016/j.compag.2018.12.006

Ruiz-Garcia, L., & Lunadei, L. (2011). The role of RFID in agriculture: Applications, limitations and challenges. *Computers and Electronics in Agriculture*, *79*(1), 42–50. doi:10.1016/j.compag.2011.08.010

Scheberl, L., Scharenbroch, B. C., Werner, L. P., Prater, J. R., & Fite, K. L. (2019). Evaluation of soil pH and soil moisture with different field sensors: Case study urban soil. *Urban Forestry & Urban Greening*, *38*, 267–279. doi:10.1016/j.ufug.2019.01.001

Yong, W., Shuaishuai, L., Li, L., Minzan, L., Ming, L., Arvanitis, K. G., ... Sigrimis, N. (2018). Smart Sensors from Ground to Cloud and Web Intelligence. *IFAC-PapersOnLine*, *51*(17), 31–38. doi:10.1016/j.ifacol.2018.08.057

Zamora-Izquierdo, M. A., Santa, J., Martínez, J. A., Martínez, V., & Skarmeta, A. F. (2019). Smart farming IoT platform based on edge and cloud computing. *Biosystems Engineering*, *177*, 4–17. doi:10.1016/j.biosystemseng.2018.10.014

Chapter 2
Image Processing Techniques Aiding Smart Agriculture

Aspira S. Tripathy
Netaji Subhas University of Technology, India

Deepak Kumar Sharma
https://orcid.org/0000-0001-6117-3464
Netaji Subhas University of Technology, India

ABSTRACT

With the ever-increasing load of satiating the agricultural demands, the transition of the orthodox methods into smart ones is inevitable. The agriculture sector for long has served as a momentous source of livelihood for many globally. It is arguably a major topic for nations of the development spectrum, contributing towards their export earnings and aiding in their GDP assessment. Thus, it is quite conspicuous that nations would work towards its expansion. In congruence, the burgeoning population and its demands have posed a threat to the environment due to extensive exploitation of resources, which in turn is escalating towards the downfall of the quality and quantity of agricultural produces requiring a 70% increment in the produces by 2050 for sustainability. To combat such hurdles, developed techniques are being employed. Through a survey of existing literature, this chapter provides a comprehensive overview of various image processing means that could come in handy for ameliorating the present scenario and shows their implied extension in the smart farming world.

INTRODUCTION

"Time is money", a beautiful quote by Benjamin Franklin, indeed lucidly highlights the importance of time. The advent and subsequent improvement of digital image processing techniques around the 1960s and 1970s allowed images to be effortlessly and effectively studied saving time, rendering a myriad of information contained within those pixels and recognizing patterns for specific purposes. However, back in time, the computing systems had constrained performance and were expensive, posing a hindrance to its proliferative usage. In parallel, machine learning was emerging as a breakthrough in the world of sci-

DOI: 10.4018/978-1-5225-9632-5.ch002

ence and technology, subsequently leading to what we coin as 'Deep Learning' today. These techniques envisioned the plausibility of machines emulating the human brain and possessing the ability to decipher and learn on its own, and therefore had potential to automate the arduous manual procedures carried out by humans, in every extant sphere. By the 2000s, fast processors and systems were widely available for which they started gaining immense popularity. Consequently, their application domain widened and researchers over the world started making use of these methodologies in order to enhance the present scenario in various sectors, including the agricultural sector, thus opening a doorway for reduction of the colossal amount of time invested on traditional agricultural practices.

Several challenges, such as deteriorating food quality, unsuitable climatic alterations, dearth of food, etc. have prevailed through the years and are escalating. As a result of recent advancements in digital technologies, image processing and machine/deep learning techniques could come handy in several ways within agricultural practices leading to the notion of "smart agriculture" or "smart farming", where digital and biotechnology, such as remote sensing, Internet of Things (IoT) and cloud computing, in conjugation with signal processing and decision-driven automating techniques alleviate the present agricultural ecosystem (Kamilaris et. al., 2017). It engulfs within its coverage soil analysis, food quality assessment, weeds recognition and many more detailed in this chapter. The traditional processes are quite arduous and time consuming, and a viable promising solution to this plight can be envisaged, buttressed by image processing and deep learning models.

Hurdles Faced and Discrepancies in Extant Agricultural Systems

To summarize, the traditional agricultural system is flawed in many ways, as is manifested through its incapability in addressing hurdles faced by agricultural practitioners:

- **Weed Management**: The extant systems require the ranchers to examine the crops fanned out over large fields intently and eradicate them by perspicacious examination. This requires a lot of time and effort, besides leaving a large fissure for human errors.
- **Cost and Time Management**: Since the farmers generally glean specific accurate information, without much reliable sources, and use manual tools, a lot of wastage occurs. Consequently, there is a good chunk of hard-earned money and time squandered.
- **Disease Prevention**: Periodic inspection and insightful knowledge is required for disease detection. This drains in a lot of expertise and time. Moreover, appropriate and best-suited recovery techniques are required for recuperation, which is usually beyond the knowledgeable domain of farmers.
- **Pest Control**: Scrutinizing the fields is a bulky task for the farmers since there is a large amount of labor and time required for the purpose. There is also a likelihood of a certain pest control medication adversely affecting an adjacent crop, thereby causing tarnish.
- **Weather and Soil Suitability**: Soil and atmospheric conditions play a crucial role in the growth of crops. Some crops require a rainy weather whereas others require a rather hot and humid weather, some require moist soil some dry. Till recently, there has been no such scheme to relate these and such overlooking has led to lesser yield.

- **Chemical Estimation**: Herbicides, fertilizers and pesticides are toxic chemicals, which need to be utilized judiciously, since a slight variation in their amounts could lead to abasement of the cultivated crop apart from emanating pollution. The traditional techniques involve spray and aerial application, inadvertently disposing off such chemicals into the environment causing serious contamination.
- **Harvest Time Estimation**: Without sensor technology, it had always been a challenging issue to estimate the precise harvest time.
- **Water Estimation and Irrigation**: Different crops require differing amounts of water for proper growth. The conventional means required irrigation through channels running across fields, without taking this into much consideration.
- **Fragmented Land Holdings**: In some countries, the regions of arable lands have depleted as an outcome of rigorous and reckless handling. This has resulted in fragmentation of suitable land sparsely. It is thus quite difficult to monitor such lands without the use of cutting-edge technological devices, such as drones and monitoring systems.
- **Gross Yield Predictions**: There has been no way of accurately predicting the gross yield traditionally. Yield estimation helps in comprehending food security and fluctuation in market prices.

It is quite pellucid that the technology has eased up the burden imposed by the above stated challenges and its allied issues, and drastically helped to contribute towards improved functioning. Concisely, we now comprehend the pool of human and materialistic resources being drained in the agricultural sector and also that a revolution was commenced with the birth of constructive technology. The prima focus of this chapter would be to provide a comprehensive review of image processing techniques for elucidate perception of the reader, through vivid exposition of their extending functionalities in aiding smart agriculture as a result of an extensive study of dedicated research. Although image processing techniques are used as feature extractors to be fed into machine/deep learning models, the corresponding models are beyond the scope of this chapter. The rest of the chapter is organized as follows. Section 2 gives an overview of generic techniques, Section 3 gives a review of study carried out in various sub-domains, followed by final conclusive remarks related to the chapter in section 4.

OVERVIEW OF GENERIC IMAGE PROCESSING TECHNIQUES

Image processing extracts its functionality from the assemblage of intermediary operations working holistically, namely (Prakash et. al., 2017):

- Image Acquisition
- Image Pre-processing
- Image Segmentation
- Image Description
- Image Classification

Various sub-sections are prevalent under each method thereby, giving us an estimate of the humongous application coverage of image processing, as shown in Figure 1.

Figure 1. Block diagram of typical stages of image processing and analysis (Adapted from (Koprowski, 2014))

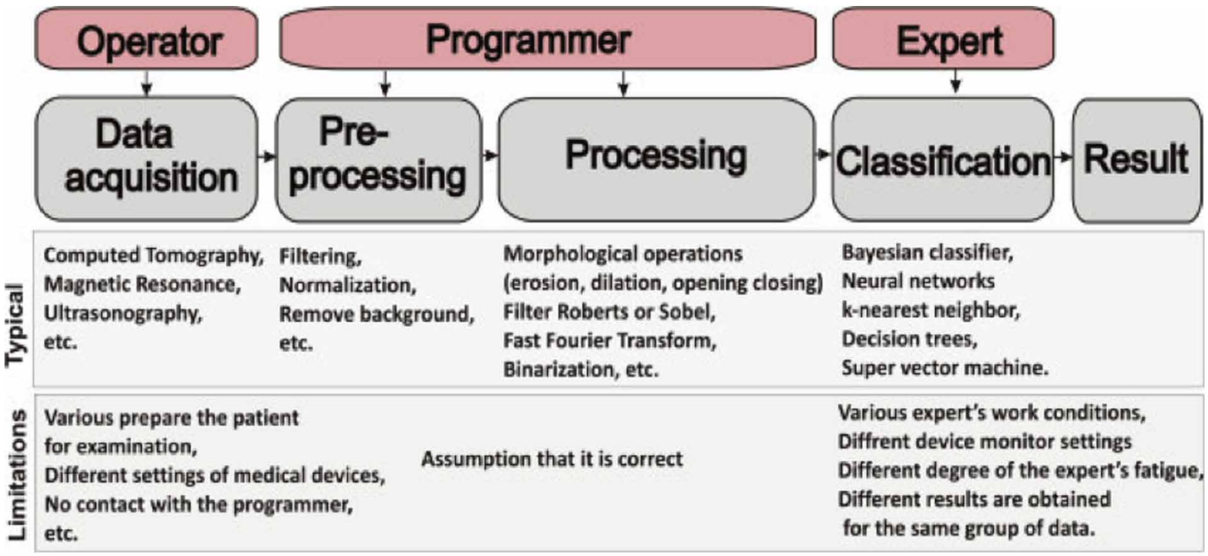


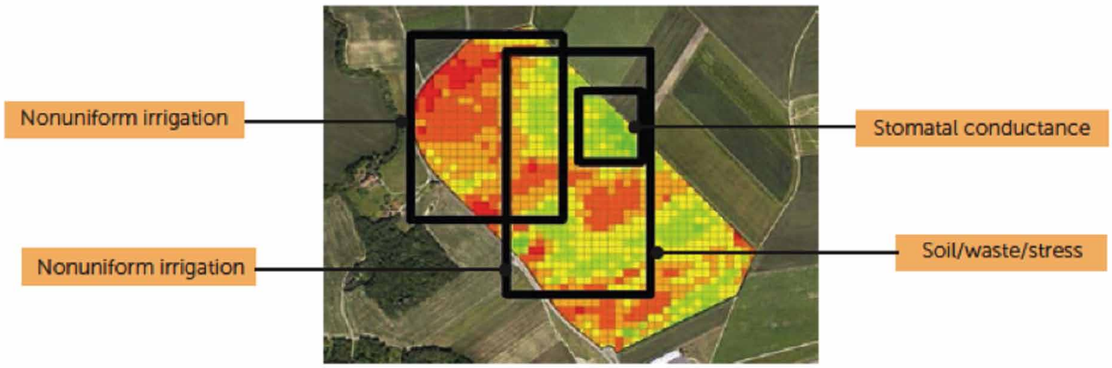
Image Acquisition and Imaging Techniques

Image acquisition is the process of data collection or actual source image collection (Prakash et. al., 2017). Differing investigations require images acquired using divergent imaging techniques like X-ray imaging, Thermal imaging, Remote Sensing, etc. (Kulalvaimozhi et. al., 2017; Koprowski, 2014; Anup et. al., 2012).

Remote sensing is basically the science of extracting crust features and gauging geo-biophysical peculiarities using electro-magnetic radiations. Microwave and optical sensors are generally in use for RS applications, such as irrigation and flood management, environment assessment, disaster monitoring and mitigation. Recently, development of drones (UAVs – Unmanned Aerial Vehicles) has led to replacement of traditional satellite-imaging methods (Udin et. al., 2016) by issuing finer control and stability. There is a probability that individual functioning sensors render imprecise, inconsistent or fragmentary information. Owing to this issue, multi-sensor image fusion is a technique of amalgamating germane data from two or more sources into one. This technique was used in enhancing the accuracy rates of vegetation classification.

Thermal imaging, a non-intrusive and a non-contact technique, does not require alterations in the surface temperature besides being capable of recording the temperature. Several studies have demonstrated that it is an apt roadway to momentous parameter selection for irrigation scheduling and studying water strain distribution. Therefore, thermal imaging has proved to be a crucial imaging technique used to correlate field saturation levels and radiation emission by recording canopy heat distribution. An example for the same is shown in figure 2. Other applications include yield forecasting, maturity grading, bruise identification, etc. Although a wide spectrum of tasks benefits from this technique, this cannot be regarded universally due to differing climatic settings and fauna physiology across regions (Kulalvaimozhi et. al., 2017; Vibhute et. al., 2012).

Figure 2. Image acquisition using thermal imaging technique (Adapted from (Roopaei et. al., 2017))



X-ray imaging has been in rampant usage for baggage inspection for import/export of illicit food products from quite a long time. However, in Haff et. al. (2008), defects and contaminants in agricultural produce were reviewed using this. Recently, X-ray micro-computed tomography (μCT) was used for characterization of food on the basis of micro-structural information obtained through image analysis (Schoeman et. al., 2016). A methodology based on digital radiography (DR) and computed tomography (CT) X-ray imaging was devised to straight-forwardly estimate the density of foods, which happens to be a structural physical property for quality evaluation. The only constraint to its burgeoning deployment across domains is its time requirement (Kelkar et. al., 2015).

Conclusively, it is evident that these techniques are in escalating phase in the agricultural sector some or the other way helping us glean relative info and forming a base for sequential processing.

Image Pre-Processing

It is a quotidian term for functions applied to images at lowest abstraction level, which can be analogously juxtaposed to mathematical normalization of dataset. The objective is a furtherance of the data that suppresses unwanted distortions and/or enhances crucial image attributes. Geometric transformations of scaling, rotation, shearing, etc., illumination corrections, morphological operations, etc. are all classified under pre-processing methods (Sonka et. al., 1993). The individual or combined application of these steps could drastically revamp the quality of feature extraction and the outcomes of respective analysis. This section gives a brief overview of the most common pre-processing techniques.

Image enhancement, a sub-category, works towards developing a vivid representation of raw data by modifying one or several intrinsic properties such that it is more susceptive to analysis. Broadly categorized into two types:

1. **Spatial Domain Methods**: Spatial domain methods involve direct pixel manipulation in image data.
2. **Frequency Domain/Transform Domain Methods**: These involve the conversion of raw image into the frequency domain prior to any manipulation with the transform coefficients. Several transforms exist such as the discrete Fourier transform (DFT), discrete Cosine transform (DCT), discrete Wavelet transform (DWT), etc.

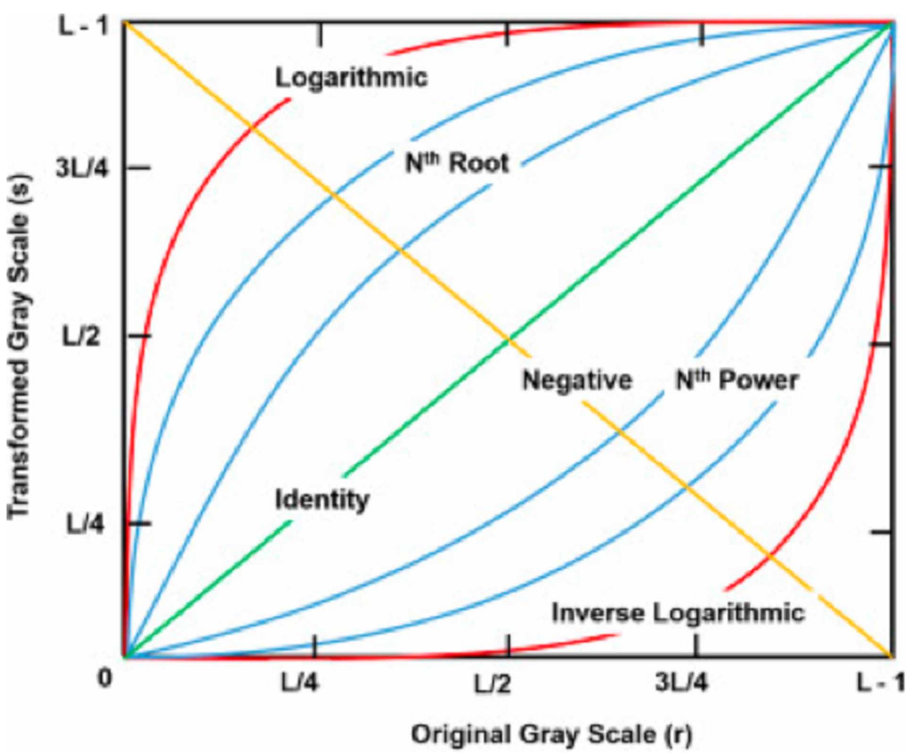
Spatial Domain Methods

Single-Point Processing

Being the simplest of all, this focuses on contrast adjustment in the image for highlighting or outlining the regions of interest. The principle approach is to define a rectangular/square neighborhood of a pixel, say *x, y,* which forms the center. It is used when the enhancement depends upon the grey level transformation of that pixel. Figure 3 shows the basic grey level transformations used for enhancement purposes. The prime types of functions are (Maini et al., 2010):

- **Linear Functions**: Negative and Identity
- **Logarithmic Functions**: Log and Inverse log
- **Powers Law Functions:** n^{th} power and n^{th} root
- Piecewise Linear transformation functions

Figure 3. Various grey level transformations (Adapted from (Gandhamal et. al., 2017))



Negative Image Formation

The simplest of all, the negative function, *N(x, y)* is given by Equation 1:

$$N(x, y) = 255 - I(x, y) \tag{1}$$

I(x, y) represents the Identity function or the image itself. This is useful in highlighting greyish or whitish details of an image amongst a darker background.

Logarithmic Transformations

Equation 2 represents the log transformation:

$$L(x, y) = c*\log(1+I(x, y)) \tag{2}$$

L(x, y) represents the logarithmically transformed output function and c is a constant, usually with a value of 1. This function focuses on mapping from narrower range to a broader range for low intensity grey values and vice versa. Basically, lighter regions undergo compression and darker regions undergo expansion.

Powers-Law Transformation (Gamma Correction)

Equation 3 is the representing equation for such functions:

$$P(x, y) = c*I(x,y)^k \tag{3}$$

k = Gamma value- ranging values and therefore result in different configurations of intensities and clarity.

Piecewise Linear Transformation

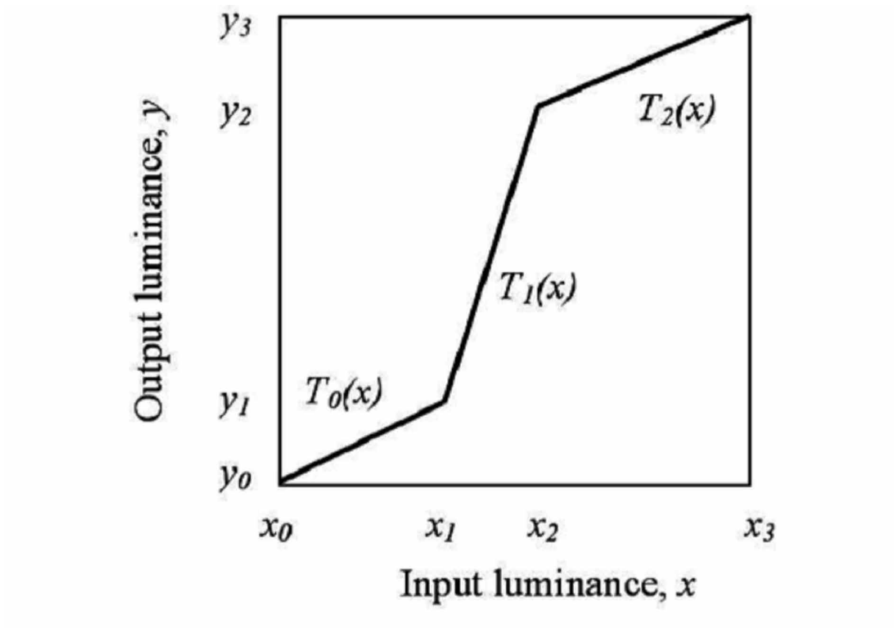
This type of user-defined and arbitrary transformation uses affine functions fragment-wise. Figure 4 gives a visual representation of such a function.

Mask Processing or Spatial Filtering or Image Smoothing

This consists of simple translation of a filter across the image, convolving the filter with the neighborhood pixels in the input image. The purpose of this image processing technique is filtering and removal of Gaussian/salt and pepper noise or spurious pixel occurrences (Ali et. al., 2014).

- **Max/Min Filter**: This filter replaces the value of the candidate pixel by the minimum or maximum pixel value.
- **Mean Filter (Neighborhood Averaging)**: The candidate pixel assumes the mean value of all the neighboring pixels.

Figure 4. Example of piecewise linear transformation function (Adapted from (Tsai et. al., 2008))



- **Median Filter**: The candidate pixel assumes the median value of all the neighboring pixels and thus is edge preserving.
- **Gaussian Filter**: It is yet another filter which leads to smoothing of the image and is comparatively faster than the median filter.
- **Gabor Filter**: The Gabor filter is an orientation-sensitive and linear band-pass filter used for detecting edges and texture analysis (Srunitha et. al., 2016).

Histogram-Based Processing

Illumination corrections, contrast adjustments and thresholding benefit from this image processing technique. A histogram highlights the grey level intensity coverage by pixels.

- Histogram Sliding
- Histogram Stretching
- Histogram Matching
- Histogram Equalization – Global and Local

Histogram Sliding

Histogram sliding refers to addition/subtraction of a constant to all the pixel values leading to a shift in the concentrated pixels towards higher grey scales (in case of addition) or lower grey scales (in case of subtraction).

Histogram Stretching

Histogram stretching is the process of multiplying/dividing the pixel values by a constant. This leads to well-balanced contrast properties.

Histogram Matching

It is basically the mapping of grey scale distribution between two different images.

Histogram Equalization

There are instances when the image is too dark with all its information skewed towards the lower side of the histogram, or too bright with all its information skewed towards the higher side of the histogram. To perceive an elucidate histogram structure, we tend to equalize the histogram by stretching the congested part or the probability density function, creating a uniformly distributed one (Anitha et. al., 2010).

Frequency Domain Methods

These techniques are dependent on orthogonal transformations having magnitude and phase as components. The former contains the frequency content of images whereas the latter is used for retrieval into spatial domain (Ali et. al., 2014). Frequency domain filtering can be majorly of two types (Naphade et. al., 1999):

1. **Low Pass Filtering**: This technique results in attenuation of high frequency components (edges) and thus a loss of overall sharpness.
2. **High Pass Filtering:** Antithetical to low pass filtering, this attenuates the lower frequency components, leading to clear edges and sharpened resultant image.

Thresholding

Thresholding is a pre-processing technique to segregate/segment various regions in an image to reveal characteristics such as background, foreground, etc. Various forms of thresholds exist:

- **Ceiling Threshold:** This allows the highest pixel intensity.
- **Floor Threshold**: This allows the lowest pixel intensity.
- **Ramp Threshold:** This is a linearly shaped threshold between the floor and the ceiling.
- **Point Threshold**: Any determinate value or point creating partitions within dataset to be analyzed.

The threshold estimation task is usually done with intent and perspicacious inspection of data. Histogram analysis is found to serve as a good pathway for this purpose. Peaks, valleys and hysteresis regions are contained within histograms, revealing a great deal of information (Krig, 2014).

Morphological Operations

Image processing has evolved over the years to circumscribe a portion of the spread-out domain of mathematical morphology within. Morphology in images is used to specify features in the form of polygon-shaped areas with defined boundaries. It entails a whole lot of set theory, topology, lattice theory, etc. for analysis of geometrical structures (JP et. al., 1994). These operations are used with the goal of eliminating structural imperfections; using a small matrix (structuring element) whose shape and size significantly bias the results (Srisha et. al., 2013). Dilation and erosion are two foundational concepts of morphology widely used.

Two definitions related to these morphological operations are fit and hit.

- **Fit**: All the pixels of the neighborhood are completely superimposed by the structuring element's pixels.
- **Hit**: An intersection/complete non-overlap of the structuring element's pixels and the corresponding neighborhood i.e. if there exists even one such pixel position where the values don't match, a hit occurs.

Refer to Figure 5 for clarification regarding the concept of dilation and erosion assuming the black dots as 255-valued pixels. In color morphology, the basic functions are MAX, MIN, and MIN-MAX. In case of MIN function, pixels crossing the defined MIN value retain their values and in case of MAX function, pixels failing to reach MAX value retain their values, while in MIN-MAX function, pixels within the specified range retain values. Let A be an image and B be the structuring element in the form of a disk or square.

Dilation

The dilation of A by B can then be defined by:

$$A \oplus \mathrm{B} = \bigcup_{b \in B} A_b$$

A_b represents the translation of A by b. In simple words, dilation results from the locus of points reached by B when the center of B traverses the entire region of A. The condition requires a hit for every pixel examined, and thus dilated images can be perceived to diffuse over a larger region.

Erosion

The erosion of A by B can then be defined by:

$$A \ominus B = \bigcup_{b \in B} A_{-b}$$

where A_{-b} represents the translation of A by –b. Simply, erosion results from the locus of points reached by the center of B while traversing inside the region contained by A. The condition requires only a fit for every pixel examined, to maintain its pixel intensity value, which becomes negative otherwise. Therefore, the images subjected to erosion can be perceived to shrink over a smaller region.

Image Segmentation

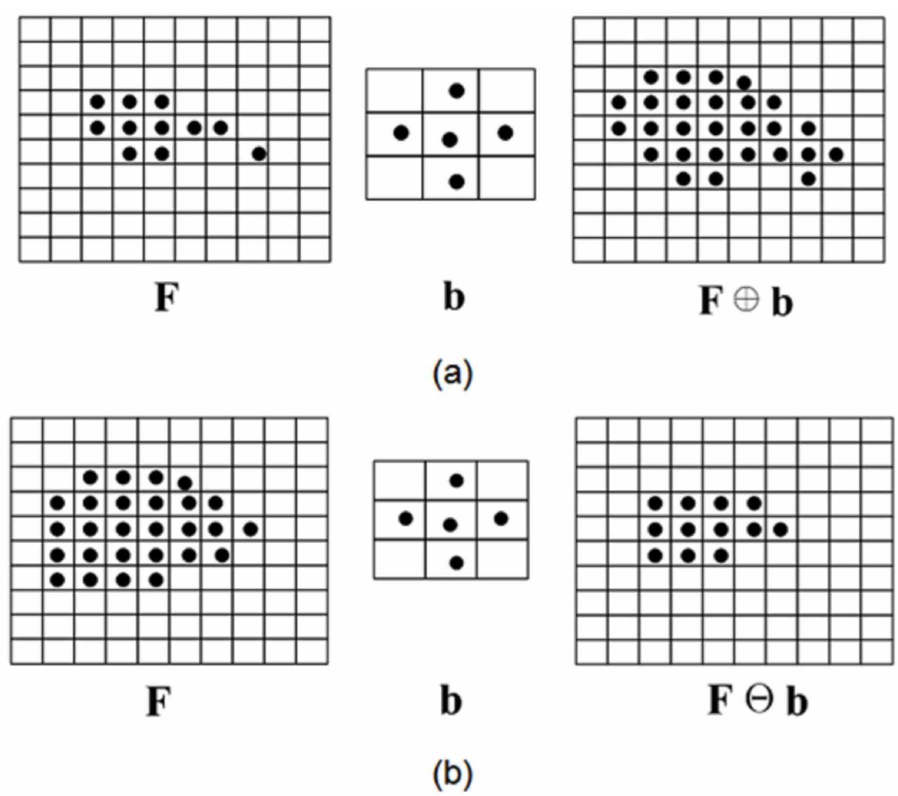
This results in the isolation of primitive entities from a given image, purpose of which is simplification of input data without the loss of apposite information. Also referred as the process of assigning labels to every pixel, it can be carried out using the thresholding techniques. This section will brief about two such techniques: Edge-based (Boundary extraction) and Region-extraction.

Edge Based Segmentation

Under this, widely used edge detectors like Sobel, Prewitt, Canny, etc. are present.

An edge is nothing but a discontinuity in the intensity grading of pixels and therefore can be identified using derivatives. This forms the radical essence of edge detection filters. Edges can be identified using first derivatives also, but their localization is difficult because of thick edge boundaries. Second

Figure 5. Morphological operations of (a) Dilation (b) Erosion (Adapted from (Li et. al., 2014))



order derivatives have a reliable response to minute details and result in zero-crossings at the peaks in first derivatives. Example- Laplacian operator for its conjugated application with Gaussian smoothing (LoG – Laplacian of Gaussian).

Hough transforms are applied over edge pixels and they are employed to pinpoint ellipses, circles, etc. obtaining sinusoidal representation in Hough space of each real point and line representation of each Hough point in real space. However, in case of irregularly shaped edges, the computation becomes expensive (Naphade et. al., 1999).

Region Based Segmentation

Watershed edge detection is based on erosion and dilation in which, image transformation takes place such that the intensity values denote the height of a three-dimensional image. A watershed is the collective set of points around the minimal intensity region and watershed lines are crest lines dividing two watersheds which apparently are the edges in the image.

Wavelet transforms are subsequently filtered obtaining high-pass and low-pass components, namely wavelet function and scaling function respectively, resulting in down-sampling of the image. Before each down-sampling, the edge details are accumulated. Haar wavelets, Meyer wavelets, etc. are some of the wavelet transforms present (Naphade et. al., 1999).

Image Description

Representation can be in the form of boundary or region: former is used when the major focus is on the shape properties whereas latter is useful for analysis of intrinsic characteristics (Prakash et. al., 2017). There are various algorithms and transforms like SIFT, SURF etc. for this.

Image Classification

This is a process of identification using classification techniques, utilizing the descriptors extracted using the above-described steps that are then pre-fed into a database to query from. The general classifiers in proliferative usage are Artificial Neural Networks (ANNs) and fuzzy logic (Prakash et. al., 2017), which are machine-learning concepts.

REVIEW OF IMAGE PROCESSING IN SMART AGRICULTURE

Smart agriculture is the inclusion of technology for the reorientation of extant systems in the agricultural sector so that unsuitable methods don't cease and hamper the productivity and revenue. Image-processing techniques, owing to their information gathering accomplishments, have engendered an army of sectors incorporating them in day-to-day manner. The following section categorizes and reviews some of the conspicuously gaining sectors and summarizes recent works within each.

LAND SUITABILITY

Soil forms the crest layers of land, which is decreasing constantly due to industrialization and population upsurge. In order to establish equilibrium between the demand and supply, it is the need of the hour to understand various soil parameters and varieties, consequently mapping befitting crops to each variety for enhanced plantation production. The mapping also succours in restricting product loss. This requires soil classification, which is nothing but the dissolution of profusely available varieties into like featured brackets. Apart from chemical properties (soil color, soil power of hydrogen (pH), soil salinity, etc. (Sereshti et. al., 2017)), physical properties such as soil moisture content, shrinkage limit, specific gravity, etc. can also be used. Soil classification can also be automated using the advanced algorithms of machine/ deep learning as in (Srunitha et. al., 2016).

Boundless plantation has led to depletion of nutrition, over the years, due to which soil preparation is an important pre-requisite to farming. In (Puno et. al., 2017), a study was demonstrated of soil nutrients and pH levels for identifying six general nutrients i.e. Nitrogen, Phosphorus, Magnesium, Calcium, Zinc and Potassium. The resultant software used the extracted mean values of HSV and CIELab spaces, as an outcome of feature extraction, after image smoothing and histogram processing followed by cropping, thresholding and masking for segmentation.

The pH, describing acidity/alkalinity, affects the nutrient availability and ultimately plant growth. Therefore, retrieval of momentous features is vital for overall improvement. These are the acid/alkaline indices (Kumar et. al., 2014) estimating fertilizers (Kamble et. al., 2017). Until recently, government labs carried out this analysis, by manual collection of samples. However, limited availability of time and labs was a generic problem faced and therefore was of very little help. Image processing has ameliorated this scenario manifold. A soil concentrated towards sourness/sweetness is inopportune for nitrogen, phosphorus and potassium absorption by the plants. Also, very alkaline soils have lesser iron and manganese. Soil color is a visually perceptive property where the numerical values of the RGB components of light reflect information about the pH content. This is based on the findings that color is correlated with the amount of organic matter, oxidation states of magnesium and iron and also the moisture content. The correlation can be quantized using the pH index, which is calculated using Equation 4:

$$pH\ \ index = \left[\frac{(Average\ Red\ component)/(Average\ Green\ component)}{(Average\ Blue\ component)} \right] \tag{4}$$

This pH index calculation was used in (Kumar et. al., 2014) where RGB values were correlated to the pH values. Soil categorization was done in (Kamble et. al., 2017) using *(4)* and software for real-time testing and analysis was delivered, using database implementation. The database was pre-fed with sample pH values and their corresponding pH index values. It provided a list of preferable crops to be cultivated on soils having different pH values, based on data collected from a reliable source.

A correlation between the physical properties and soil type was presented in (Karisiddappa et. al., 2010). Fractal Dimension, a mathematical descriptor, was calculated using the Box Counting Method after binarization of captured image through thresholding. The box counting method defines fractal dimension by Equation 5:

$$FD = \frac{\log\left(N\left(s\right)\right)}{\log\left(\frac{1}{s}\right)} \tag{5}$$

where N(S) = number of one's in a binary image, circumscribed by a square box of edge length 's'. The physical properties and FDs were correlated using regression models in the form of graphs, which were then referred for an average FD of test images. The same techniques as in (Karisiddappa et. al., 2010) and (Kumar et. al., 2014) were used to determine the chemical and physical properties of soil in (ManikandaBabu et. al., 2016).

Soil porosity and surface-soil cracks are two other concepts of great significance in agriculture. The quantification of pore sizes or the grain size distribution in soil could give us an idea about the rates of solute transport, water circulation, etc. (Naphade et. al., 1999) whereas the geometric calculations of surface soil cracks can intimidate about the hydraulics and mechanics of soil (Shit et. al., 2015). Conventionally, intrusion techniques and visual assessment methodologies were applied respectively, but with the emergence of useful technology, accurate measurements concerning the area, width, etc. could be taken. Seasonal deviations appear to be a common reason for soil cracks, but continual climatic fluctuations could drastically exacerbate the situation. Kumar et. al. in (Shit et. al., 2015) studied the cracking process and used techniques of smoothing, binarization, thresholding and erosion/dilation for the purpose.

Apart from soil properties, classification can be done on the basis of diversified soil-types present such as clay, loam, peat, etc. Traditionally, it was done by drilling boreholes for samples. A comparatively feasible investigation alternative, known as Cone Penetration Testing (CPT), appeared later, which involved a metallic cone being penetrated through the study region in order to estimate soil strength. Both these techniques drained a lot of manual labor and cost. Image-processing aided classification is tremendously inexpensive in comparison, besides being available on digital systems even in remote areas. Authors used image analysis in (Srunitha et. al., 2016) for binary classification between sandy and non-sandy soils, three-class classification among sand, clay and peat and a seven-class classification. Database was employed and the training and query images were subjected to a series of image processing techniques. Gabor filter, low mask filter, histogram processing and color quantization were executed under image pre-processing and statistical measures of mean and standard deviation were employed to study the color, texture, and shape features. In 2018, soil of Rajasthan, India was categorized using noise removal, HSV space conversion, histogram processing and sharpening followed by segmentation and classification (Hement et. al., 2018).

DISEASE DETECTION

Fungal/leaf/parasitic infections etc. not only affect productions severely by plaguing healthy crops, but also the financial source of agriculturists and human health. Hence, detection of diseases at early stages is more crucial to the entire community than said. Infestation hampers the yield and quality of the produces. In addition, excessive utilization of disease-specific chemicals often further deteriorates the quality. Therefore, monitoring infections and detecting them beforehand is important for effective management and planning curing measures so that agricultural standards are sustained. The nascent infective growth is majorly microscopic in visual range and thus the diagnosis phase is constrained by human vision. Although diseases are caused by numerous pathogens like nematodes, bacteria, etc., fungi are the rifest disease-infecting pathogen. Disease study can be summarized as the investigation of perceptible patterns/peculiar trends depending upon visual symptoms such as spots/lesions and their numeric estimation, size or color that may give an estimate of plant mortification. Adding to visual limitations, manual inspection is too arduous such that humanly issues of fatigue, subjectivity, etc. might as well grant access to man-made mistakes into the investigation. Therefore, automated systems are a boon for such analysis, curbing out the probability of man-errors.

Disease detection in apples and grapes was investigated in Jhuria et. al. (2013) using the set of color, texture and morphological features. Study suggested that highest accuracy was by morphological features. HSI was used due to RGB's receptiveness towards varying illumination and positions. Erosion was used and for texture feature, Daubechies 2-dimensional wavelet packet decomposition was used. Authors in Pujari et. al. (2015) primarily focused on fungal infections in fruit, vegetable, cereal and commercial crops. For fruit crops, thresholding, median filters, watershed and canny edge detectors, GLCM (Grey Level Co-occurrence Matrix) and GLRLM (Grey Level Run Length Matrix) were used and Euclidean distance was conceptualized for classification. For commercial crops, DWT (Discrete Wavelet Transform) was used as a feature extractor. For cereal crops, RGB and HSI spaces were used and affected regions were first identified using Radon transformation and projection algorithm. Local Binary Patterns (LBP) were utilized for vegetable crops. In (Rau et. al., 2017), raspberry pi was included in an IoT framework to provide real-time transmission of data to farmers for disease detection in paddy leaves. A database was designed with different diseases labeled in terms of physical properties like homogeneity, entropy, etc.

Dhole et. al. (2016) reviewed unhealthy leaves for citrus fruits for which space conversion into YCbCr system and segmentation, followed by shape, texture and colour feature extraction was done. Edge and colour histograms were used in Tajane et al. (2014) for identifying leaf features in medicinal plants. Edge histogram is simple histogram plotting of the prominent edges found by canny edge detector. Husin et. al. (2012) processed chilly leaves for health status monitoring of the plant. Fourier-filters, morphology and edge detectors were deployed in pre-processing. In Khirade et al. (2015), image segmentation using Boundary and spot detection algorithm, K-means clustering, and Otsu's Thresholding was done besides histogram equalization as in Thangadurai et al. (2014). The foremost required space conversion into HSI whereas Otsu's was applied using RGB. In Otsu's, the colour pigments affected, that led to the greenness factor, served as a reliable parameter. Features were extracted using GLCM and hue/brightness components of HSI and CIELab. Authors in (Dey et. al., 2016) also assessed leaf rot diseases in Betel Vine using Otsu's and considered size and complexity reduction for faster analysis. Hue component of HSV accurately described the rotten area, which was reassured by histogram processing prior to subjecting it to Otsu's threshold-based segmentation. Tanvi et. al. in 2016 (Mehra et. al., 2016) studied Septoria Leaf spots on tomato leaves using RGB space.

Noise suppression and smoothing have emerged as important pre-processing steps in nearly all reviewed papers. Therefore, noise-eradicating filters need to be studied with efficacy like in Valliammai et al. (2012). Two major noise variants, Gaussian and speckle noise were found guilty of blurring leaf veins and distorting the shape/size of leaves. In order to eliminate both, a hybrid noise removal method was formulated in (Mishra et. al., 2017). Other techniques like histogram equalization and median filtering can widely be exploited for enhancement purposes whereas Gabor filtering, Wavelet transformation, co-occurrence methods, etc. can be widely used for extraction purposes (Dhole et. al., 2016). It is pellucid how technology has modified the traditional systems and the concern shifts to the incorporation of these frameworks into systems for real-time application. Several erudite researchers have tried this using android/web bases. In 2015, Bhange et al. (2015) proposed a web-based tool for fruit disease detection focusing on pomegranates infested by Bacterial blight. Image-resizing with colour, morphology and CCV features were extracted. Colour features used 3-bin histogram for RGB components of the image. Morphological analysis included erosion, which rendered image boundary by subtraction between original and eroded images. CCV (Colour Coherence Vector) is a histogram-based method dealing with pixel characterization into coherent and non-coherent ones. Image blurring and discretization of colour values are pre-requisite for extracting these features. Dixit et. al. in 2018 (Gajanan et. al., 2018) subsumed SURF descriptors/locators in disease detection analysis for android systems. SURF is divided into three algorithmic steps: (1) Scope Point Detector, (2) Local Surrounding Descriptor, (3) Matching. After SURF, blob analysis was carried following pattern matching mechanism juxtaposing original and test images on the basis of four features: colour (HSI), morphology (erosion), texture (wavelets) and hole structure. Astonkar et. al. (2018) developed software, using image pre-processing, primarily scaling for accelerating the process, and image segmentation using K-means clustering. In Raichaudhuri et al. (2017), they provided a system for detection of wheat leaf diseases using canny edge detectors, k-means algorithm and GLCM features.

YIELD AND FERTILIZER ESTIMATION

An important consequence of advancements is the recent enlightenment on the possibility of yield-estimation. Yield (counting factor for crop-productivity) gives guidance on managing market/field variables, estimating area-specific chemical usage and product-price. A major benefiter of this is precision-agriculture: a sub-domain of smart-agriculture. It refers to the early estimation of crop, for pre-planning at each manufacturing stage, for efficient workflow. Traditionally, fertilizer application, liming, etc. were decided by taking the climate into consideration and not the plant characteristics individually. With precision-agriculture it became possible to track plant-peculiarities and therefore plan fertilizer-application efficiently. Not only is production forecasting of utter significance for individual agriculturists, but it also has a straightforward impact on the agricultural management policies. Leaving to various hurdles, in majority cases, only a limited portion of the land was examined and the rest of the data was extrapolated, thus leading to inaccurate results and conclusions by a great margin. The image processing frameworks can help ease the burden by incorporation in UGVs (Unmanned ground vehicles) for farm investigation. With the increased imbibing of image processing into the various sectors, yield-estimation practices have been worked on quite a lot over the past years, simplifying the process to a commendable extent. However, the major hurdle faced by detection systems is the variability of farm-image data due to illumination alterations, occlusions, shadowing effects, etc. (Bargoti et. al., 2016).

There are a large number of fruits whose detection algorithms included image processing like grapes (Font et. al., 2015; Nuske et. al., 2014), apples (Kim et. al., 2015; Silwal et. al., 2014), mangoes (Payne et. al., 2013; Stein et. al., 2016), etc. Colour information was used in Ramos et. al. (2017) and Avendano et al. (2017) to identify coffee branches and estimate fruits. In the latter, the detection was carried out by video processing through mobiles. Colour and shape features were used in Wang et al. (2012) for counting the number of apples in the orchards. Red apples were identified using thresholding in HSV whereas green apples were identified using specular reflection with low saturation and high brightness. Hung et al. (2015) proposed a generalized multi-scale feature technique to multi-class segmentation for fruit estimation in apple trees. Bargoti et al. (2016) focused on image segmentation technique for yield estimation in apple orchards. This research was a sequel of their previous one, in which one of the network architectures had been defined. It further extended into CNNs (Convolutional Neural Networks) and proved the usefulness of inclusion of metadata (contextual information) in systems for relatively higher accuracy. The watershed techniques and circular Hough transform algorithms were employed out of which the former gave outshining results. According to Longsheng et. al., there had been no study on kiwi fruit estimation until their attempt in 2018. Therefore, they proposed an android mobile phone (AMP) based kiwi-estimation system using image-processing tools. The colour information was analyzed in RGB, HSV and L*a*b* spaces and the distribution was scrutinized using the kurtosis value (Fu et. al., 2018) and normalized after appropriate colour channel and threshold value selection. By evidence, the kurtosis values of a* and H components were the highest. Noise was removed using erosion and dilation and canny edge detector was used. Then, contrast adjustments using histogram equalization was done followed by Otsu's thresholding and bitwise AND operation. The bitwise AND operation was solely applied for speeding up purpose. The fruit number was estimated by recognizing the fruit calyxes in the final processed image. A reference board, hung under the canopy, had Sobel detector in use for edge detection of reference board. Ultimately, the fruit density was calculated using the estimated pixel-area and reference board pixel-area using an equation. Fruit density and orchard area on multiplication gave an estimate of the number of fruits.

Citrus products' forecasting prior to harvesting is very crucial due to similar reasons. The authors of Dorj et al. (2017) used histogram-based thresholding for segmentation due to their lesser requirements of execution. The major limitation posed to image acquisition was illumination, since direct sunlight lead to saturation and shadows caused fragmentation while segmentation. Thus pictures were taken under diffusive lighting. RGB to HSV conversion, thresholding, smoothing, watershed segmentation (which also segregated connected citrus') and ultimately fruit counting using blob detection, were the respective major steps taken. For noise removal purpose, edge preserving median filter was used. Two kinds of watershed segmentation were used in this paper: (1) Distance Transform Watershed Segmentation (2) Marker-Controlled Watershed Segmentation. In (Maldonado et. al., 2016), bas-relief representation was emphasized for its vivid simulation of texture and depth features in an image. The bas representation highlighted the fruits by demarcating highlighted regions, into upper and lower parts. The ratio of these two parts served as a spatial mask for fruit estimation, which proved to have a higher value in case of fruit. For acquiring such a representation, a sequence of image processing techniques had to be used. Single-point log transformation and grey-scale histogram equalization were applied to the V component of HSV converted image. The reason for selecting the V component and respective discarding of the other two was because of the brightness information provided by it. However, soil was segregated by thresholding the H component of HSV representation. Histogram stretching was employed in addition to histogram equalization for contrast adjustments and illumination standardization. After a complete application of the above-mentioned techniques, Sobel operator in the horizontal direction and Laplace operators were used for texture analysis. The final processed image was subjected to Hough circle detector to render the final outcome of yield prediction.

Apart from edible substances' production, there are other industries that benefit from yield-estimation practices such as the turf-grass sod industry. This industry has an escalating reputation due to the wide-ranging benefits of turf-grass in urban landscapes, like aesthetic and recreational ones. Excessive nitrogen rates can retard root development and therefore, fertilization is an important task to distribute apt amounts of nitrogen to varying sites. Fertilizer estimation can also be carried out as a result of crop yield estimation since it has been long from when variable rate fertilization (VRF) is an issue in the sub-domain of precision agriculture. For example, in (Chung et. al., 2018), the grass-growth status is monitored to estimate the variable site-specific application of fertilizers. Excess green index (ExG) was utilized for binarization and segmentation after which an inverse linear relationship between grass growth (GG) and fertilizer recommendation (FR) was proposed.

PEST DENSITY ESTIMATION AND MANAGEMENT

Integrated pest management plays vital role in diminishing production-loss and chemical-usage. Farmers have ranked pests to be one of the top issues faced whether an effect on the stored grain stocks or live infestation on plants. To overcome yield losses, they resort to pesticides/insecticides posing a hazard to humans/animals/environment and the crops themselves. The traditional means included spraying-technology dependent upon the climate more than the root-cause. Moreover, it was labourious and time draining. Thus, there is a requirement for proper analysis before chemical application to altogether ameliorate the efficacy of the prevalent agronomic practices. This requirement has led to the development of various technologies, some of which have already been standardized for commercial applications. These specifically refer to integrated pest management.

In 2016, Maharlooei et al. (2017) developed an algorithm for soybean-aphid density estimation using image processing. Two existing techniques were discussed: (1) Pan trapping (2) Speed scouting. The log-transformed pan count was strongly correlated to the holistic plant counts, which formed the basis of their usage. Both these techniques required periodic sampling. They initiated with manual cropping and segmentation of the leaf-area. ExG, followed by hole filling and image-reconstruction, was done to obtain final segmented image with background elimination. ExG was implemented using the equation: 2*G-R-B, intensifying greens and 0.3-thresholding was done for binarization. Later, interest area was isolated from leaf-background, in the HSI domain then smoothened and enhanced using the hue/intensity components for specific localization. Finally, the distinction between aphids, yellow spots and dead skin was done on the basis of surface-area variation. Authors in Jige et al. (2017) estimated the whitefly density on cotton leaves. Initially, the image was converted into HSV for easier perception. Then background elimination and adaptive thresholding were carried out. The reason for using adaptive thresholding: variations in illumination and generic environmental conditions affected the threshold. After this step, whiteflies were visible with red circles round the overlapping whiteflies. Erosion/dilation were used for separation of overlapping cases.

There are several other researches focusing on whitefly detection: In Ghods et. al. (2016), a hypothesis system was first designed for segmentation purposes after grey-scale conversion and then colour, shape (morphological opening) and texture (entropy) features were extracted. In Barbedo et. al. (2014), whiteflies were identified and counted on soybean leaves. Their algorithm not only detected full-growns but also nymph-staged whiteflies. In Huddar et. al., (2012), relative pixel difference was calculated and thresholded for segmentation purposes, after YCbCr conversion. In Brinda et al. (2016), three feature extraction techniques were juxtaposed: histogram-oriented gradient, Gaussian mixture, Gabor filtering. Very recently in Sun et al. (2017), whiteflies and thrips were counted on sticky traps – which are indicators of pest density, using 2-D Fourier transform. It is an early researched fact that the count of pests on the upper edge and sides of the traps gives idea about the total population density. However, it is strongly dependent on human expertise and memorizing capabilities, thus leading to inaccurate outcomes very often. The 2-dimensional Fourier-transform was thus utilized, for automation, in the calculation of a pest density-estimating index given by Equation 6:

$$Pest_{index} = \frac{S_0(0,0) - S_n(0,0)}{S_0(0,0)} \tag{6}$$

$S_0(0,0)$ being the mean amplitude in Fourier spectrum, with sticky trap without any pests and $S_n(0,0)$ being the mean amplitude in Fourier spectrum with *n* pests trapped on the sticky trap. This process put forward by the authors was time-efficient due to non-inclusion of boundary/geometry characteristics. In Martin et. al. (2015), extended region grow algorithm was used to estimate pest population on paddy fields. The region growing methods divided the sample space into regions with properties similar to the ones described. Pavithra N. et. al. in (2017) compared RGB, YCbCr and HSV colour models only to find outstanding performance by HSV model. Manual and Otsu's thresholding was also compared with manual method obtaining an upper hand. The final image was obtained with whiteflies detected as white pixels over black background pixels after noise removal through dilation/erosion operations. Zhu et. al. (2018) in 2018 estimated pest population in stored grains and developed a mobile-embedded system for the same. Stress was paid on insect contour through colour space transformation, binary processing, boundary tracing, etc. Two methods of pest counting were discussed: Template matching method and connected domain-based histogram. The former focused on body postures requiring database matching (inefficient for insects having similar body characteristics and more complex in case of increasing test samples) whereas in the latter, a connected domain with 18 to 24 pixels was categorized as one insect.

SMART IRRIGATION

The yield and quality depend on a number of varying factors, optimum water availability being one of them. Smart agriculture also encompasses uniform and periodic water dissemination across fields. As stated in Barkunan et al. (2019), agricultural sector uses 85% of freshwater resources globally available for cultivation. This, in turn, affects the water availability levels for quotidian chores of the burgeoning population. Farmers are still resorting to manual irrigation that adds to the chance of over/under watering. The rudimentary technique of irrigation being practiced since the ancient ages is the method of flood irrigation which releases water into the entire farm until covered, leading to drastic wastage of water resources, soil erosion, plant spoilage, etc. An early alternative to this drawback was drip irrigation (micro-irrigation), using pressurized pipelines, valve systems and specialized equipment such as drippers, for optimizing the water usage by directly letting the water seep drop wise through the root-zone/soil-crest. But, a large species of flora do not require the same amount of water throughout their growth span, for example, a generic pattern is the requirement of mere 25% in the early stages and 50% in the mid-ages. Also, it is a lucid observation that there has been a gradual decrement in the groundwater levels over the past decade, in addition to precarious monsoon season patterns. This necessitates the irrigation automation using economical/preserving-demeanour towards water resources to avoid exhaustion conditions in the near future and guarantee increased yield.

In Rau et. al. (2017), the authors researched on a weather-based smart-irrigation system over IoT network inclusive of Raspberry-pi, sensing-technology, solenoid-valve and dongle. The readings from the DH11 (temperature/humidity sensor) were transmitted to the raspberry-pi, which was in charge of the opening/blocking operations for handling watering intervals/amounts. The flow threshold had been fixed in accordance to the water loss, estimated through the Penman-Monteith Evapotranspiration formula. In (Car et. al., 2012), an advantageous benefit of using SMS interface for balance-based irrigation scheduling DSS (Decision Support systems) was found. The water balance was maintained and updated by reverse messaging of drip-runtime information and rainfall data by the farmers. Authors in Barkunan et. al. (2019) integrated sensors based on temperature, humidity, illumination-intensity, rain-level and wetness/water-content of soil, for paddy cultivation. The smart sensor was inclusive of ARM microcontroller, GSM module, sensor unit, smart-phones and motor-control unit. A system was designed for determining the soil water content on the basis of simple image processing. The image data acquired through the smartphone was first converted into grey-scale using the equation in (ITUR, 2011) given by: Grey-scale image = 0.2989R + 0.5870G + 0.1140B and then subjected to histogram analysis. The decision that the soil is wet and didn't require to be irrigated was left over to the condition of the total number of grey-scale image pixels. If it exceeded over 5000 around pixel intensity of 200, the soil was considered wet enough to not be irrigated. In other cases, the system responded differently according to pre-defined thresholds. The authors of Gutiérrez et. al. (2014) placed smart sensors in the root zone of crops for soil-moisture content and temperature estimation, which in the form of wireless data was transmitted via public mobile network to the web-based server.

CONCLUSION

In this chapter, the main focus is towards image-processing techniques and how they aid in achieving sustainable and enhanced agriculture. The emergence of smart technology has been explained in detail for a clear understanding in the initial part of the chapter. The constituent sections within the broad domain of image processing are stated and explained along with the inclusive generic methodologies. Through the years, a variety of techniques have been developed for analysis and this chapter has covered different patches extending. Smart agriculture, being a major benefiter of such techniques has been researched quite a lot through the years. It being a revolutionary evolving concept itself, is consisting of a myriad of sub-domains. While this chapter hasn't carried out an exhaustive overview, it has tried including the momentous sub-domains. The in-depth review of the scholarly work on these sub-domains from the past decade has been vividly presented and summarized, including the various image-processing techniques used, some of which have been explained in the former sections. This way, the chapter concisely summarizes the extant image processing mechanisms for further research on the same lines or extended research into other domains.

REFERENCES

Ali, Sarah. (2014). Image Enhancement - Spatial vs. Frequency Domain Filters. . doi:10.13140/2.1.1256.7042

Ammad-udin, M., Mansour, A., Le Jeune, D., Aggoune, E. H. M., & Ayaz, M. (2016). *"UAV routing protocol for crop health management," 2016 24th European Signal Processing Conference* (pp. 1818–1822). Budapest: EUSIPCO; doi:10.1109/EUSIPCO.2016.7760562.

Astonkar, S. R., & Shandilya, V. K. (2018). Detection and analysis of plant diseases using image processing technique. International Research Journal of Engineering and Technology (IRJET), 05(04), 4. Retrieved from https://www.irjet.net/archives/V5/i4/IRJET-V5I4717.pdf

Avendano, J., Ramos, P. J., & Prieto, F. A. (2017). A system for classifying vegetative structures on coffee branches based on videos recorded in the field by a mobile device. Expert Systems with Applications, 88, 178e192.

Barbedo, J. (2014). Using Digital Image Processing for Counting Whiteflies on Soybean Leaves. *Journal of Asia-Pacific Entomology*, *17*(4), 685–694. doi:10.1016/j.aspen.2014.06.014

Bargoti, S., & Underwood, P. James. (2016). Image Segmentation for Fruit Detection and Yield Estimation in Apple Orchards. *Journal of Field Robotics*. doi:10.1002/rob.21699

Barkunan, S. R., Bhanumathi, V., & Sethuram, J. Smart sensor for automatic drip irrigation system for paddy cultivation, Computers & Electrical Engineering, Volume 73, 2019, Pages 180-193, ISSN 0045-7906, doi:10.1016/j.compeleceng.2018.11.013

Bhange, M., & Ali, H. (2015). Smart Farming: Pomegranate Disease Detection Using Image Processing. *Procedia Computer Science*, *58*, 280–288. doi:10.1016/j.procs.2015.08.022

P. Brinda and Dr.M.Pushparani, "Analysis of Early Leave Pest Detection", IJRITCC, May 2016.

Car, N., Christen, E., Hornbuckle, J., & Moore, G. (2012). Using a mobile phone Short Messaging Service (SMS) for irrigation scheduling in Australia – Farmers' participation and utility evaluation. *Computers and Electronics in Agriculture*, *84*, 132–143. doi:10.1016/j.compag.2012.03.003

S.O. Chung, M.S.N. Kabir, Y.J. Kim,Variable Fertilizer Recommendation by Image-based Grass Growth Status,IFAC-PapersOnLine,Volume 51, Issue 17,2018,Pages 10-13,ISSN 405-8963,. doi:10.1016/j.ifacol.2018.08.053

Dey, K., Sharma, M., & Meshram, M. R. Image processing based leaf rot disease, detection of betel vine (Piper BetleL.), in: Proceedings of the International Conference on Computational Modeling and Security (CMS), 2016, pp. 748–754. doi:10.1016/j.procs.2016.05.262

Dhole, S. A., & Shaikh, R. P. (2016). Review of leaf unhealthy region detection using image processing techniques. *Bulletin of Electrical Engineering and Informatics*, *5*(4), 3. doi:10.11591/eei.v5i4.498

Dorj, U.-O., Lee, M., & Yun, S. (2017). An yield estimation in citrus orchards via fruit detection and counting using image processing. *Computers and Electronics in Agriculture*, *140*, 103–112. doi:10.1016/j.compag.2017.05.019

Dr.K.Thangadurai, K.Padmavathi, "Computer Vision image Enhancement For Plant Leaves Disease Detection", 2014 World Congress on Computing and Communication Technologies.

Font, D., Tresanchez, M., Mart'ınez, D., Moreno, J., Clotet, E., & Palac'ın, J. (2015). Vineyard yield estimation based on the analysis of high resolution images obtained with artificial illumination at night. *Sensors (Basel)*, *15*(4), 8284–8301. doi:10.3390/s150408284 PubMed

Fu, Longsheng & Liu, Zhihao & Majeed, Yaqoob & Cui, Yongjie. (2018). Kiwifruit yield estimation using image processing by an Android mobile phone. 51. 185-190. . doi:10.1016/j.ifacol.2018.08.137

Gajanan, D.E., Shankar, G., & Keshav, G.V. (2018). Detection of Leaf Disease Using Feature Extraction for Android Based System.

Gandhamal, A., Talbar, S., Gajre, S., Hani, M., Fadzil, A., & Kumar, D. (2017). Local Gray Level S-Curve Transformation – A Generalized Contrast Enhancement Technique for Medical Images. *Computers in Biology and Medicine*, *83*, 120–133. doi:10.1016/j.compbiomed.2017.03.001 PubMed

Gutiérrez, J., Villa-Medina, J. F., Nieto-Garibay, A., & Porta-Gándara, M. Á. (2014, January). Automated Irrigation System Using a Wireless Sensor Network and GPRS Module. *IEEE Transactions on Instrumentation and Measurement*, *63*(1), 166–176. doi:10.1109/TIM.2013.2276487

Haff, R. P., & Toyofuku, N. (2008). X-ray detection of defects and contaminants in the food industry. *Sensing and Instrumentation for Food Quality and Safety*, *2*(4), 262–273. doi:10.1007/s11694-008-9059-8

Hung, C., Underwood, J., Nieto, J., & Sukkarieh, S. (2015). A feature learning based approach for automated fruit yield estimation. In *Field and Service Robotics (FSR)* (pp. 485–498). Brisbane, Australia: Springer; doi:10.1007/978-3-319-07488-7_33.

Jhuria, M., Kumar, A. P., & Borse, R. (2013). Image processing for smart farming: Detection of disease and fruit grading. 2013 IEEE Second International Conference on Image Information Processing (ICIIP-2013), 521-526. doi:10.1109/ICIIP.2013.6707647

N. Jige, Monica & R. Ratnaparkhe, Varsha. (2017). Population estimation of whitefly for cotton plant using image processing approach. 487-491. . doi:10.1109/RTEICT.2017.8256644

JP, Serra & Soille, P. (1994). Mathematical Morphology and Its Applications to Image Processing (Computational Imaging and Vision). . doi:10.1007/978-94-011-1040-2

Jr, M. (2016). Automatic green fruit counting in orange trees using digital images. *Computers and Electronics in Agriculture*, *127*, 572–581. doi:10.1016/j.compag.2016.07.023

Kamble, U., Shingne, P., Kankrayane, R., Somkuwar, S., & Kamble, S. (2017). Testing of agriculture soil by digital image processing. IJSRD-International Journal for Scientific Research & Development, 5(01), 3. Retrieved from http://ijsrd.com/Article.php?manuscript=IJSRDV5I10396

Kamilaris, A., Kartakoullis, A., & Prenafeta-Boldu, F. X. (2017). A review on the practice of big data analysis in agriculture. *Computers and Electronics in Agriculture*, *143*, 23–37. doi:10.1016/j.compag.2017.09.037

Karisiddappa, Ramegowda, & Shridhara, S. (2010). Soil characterization based on digital image analysis. In Indian Geotechnical Conference- GEOtrendz (p. 4). Retrieved from https://gndec.ac.in/~igs/ldh/conf/2010/articles/017.pdf

Kelkar, S., Boushey, C. J., & Okos, M. (2015). A Method to Determine the Density of Foods using X-ray Imaging. *Journal of Food Engineering*, *159*, 36–41. doi:10.1016/j.jfoodeng.2015.03.012 PubMed

Khirade, D. Sachin & Patil, A.B. (2015). Plant Disease Detection Using Image Processing. Proceedings - 1st International Conference on Computing, Communication, Control and Automation, ICCUBEA 2015. 768-771. 10.1109/ICCUBEA.2015.153

Kim, D., Choi, H., Choi, J., Yoo, S. J., & Han, D. (2015). A novel red apple detection algorithm based on AdaBoost learning. IEIE Transactions on Smart Processing & Computing, 4(4), 265–271. doi:10.5573/IEIESPC.2015.4.4.265

Koprowski, R. (2014). Quantitative assessment of the impact of biomedical image acquisition on the results obtained from image analysis and processing. *Biomedical Engineering Online*, *13*(1), 93. doi:10.1186/1475-925X-13-93 PubMed

Krig, S. (2014). Image Pre-Processing. In Computer Vision Metrics. Berkeley, CA: Apress. doi:10.1007/978-1-4302-5930-5_2

Kulalvaimozhi, V. P., Alex, M. G., & Peter, S. J. (2017). Image processing in agriculture. International Journal Of Advancement In Engineering Technology, Management and Applied Science (IJAETMAS), 04(03), 10. Retrieved from http://www.ijaetmas.com/wp-content/uploads/2017/04/IJ17M0314.pdf

Kumar, V., Vimal, B., Kumar, R., Kumar, R., & Kumar, M. (2014). Determination of soil pH by using digital image processing technique. *Journal of Applied and Natural Science*, *6*(1), 14–18. doi:10.31018/jans.v6i1.368

Li, W.-L., Xie, H., Li, Q.-D., Zhou, L.-P., & Yin, Z.-P. (2014). Section Curve Reconstruction and Mean-Camber Curve Extraction of a Point-Sampled Blade Surface. *PLoS One*, *9*(12), e115471. doi:10.1371/journal.pone.0115471 PubMed

Maharlooei, M., Sivarajan, S., Bajwa, S., Harmon, J., & Nowatzki, J. (2017). Detection of soybean aphids in a greenhouse using an image processing technique. *Computers and Electronics in Agriculture*, *132*, 63–70. doi:10.1016/j.compag.2016.11.019

Maini, R., & Aggarwal, H. (2010). A Comprehensive Review of Image Enhancement Techniques. *Journal of Computers*, 2.

ManikandaBabu. C. S., & Arun Pandian, M. (2016). Determination of physical and chemical characteristics of soil using digital image processing. International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE), 20(2), 5. Retrieved from http://www.ijetcse.com/wp-content/plugins/ijetcse/file/upload/docx/9ICAEECI-017-pdf.pdf

Martin, A., Sathish, D., Balachander, C., Hariprasath, T., & Krishnamoorthi, G. (2015). Identification and counting of pests using extended region grow algorithm. 2015 2nd International Conference on Electronics and Communication Systems (ICECS), 1229-1234.

Mehra, T., Kumar, V., & Gupta, P. "Maturity and disease detection in tomato using computer vision," 2016 Fourth International Conference on Parallel, Distributed and Grid Computing (PDGC), Waknaghat, 2016, pp. 399-403.doi:10.1109/PDGC.2016.7913228

Mishra, B., Nema, S., Lambert, M., & Nema, S. "Recent technologies of leaf disease detection using image processing approach 2014; A review," *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, Coimbatore, 2017, pp. 1-5. doi: 10.1109/ICIIECS.2017.8275901

Naphade, K. S. "Soil characterization using digital image processing" (1999). Theses and Dissertations. 630. https://preserve.lehigh.edu/etd/630

Nuske, S., Wilshusen, K., Achar, S., Yoder, L., & Singh, S. (2014). Automated visual yield estimation in vineyards. *Journal of Field Robotics*, *31*(5), 837–860. doi:10.1002/rob.21541

N. Pavithra and V. Murthy, "An Image Processing algorithm for pest detection", pices, vol. 1, no. 3, pp. 24-26, Jul. 2017.

Payne, A. B., Walsh, K. B., Subedi, P. P., & Jarvis, D. (2013). Estimation of mango crop yield using image analysis – segmentation method. *Computers and Electronics in Agriculture*, *91*, 57–64. doi:10.1016/j.compag.2012.11.009

Prakash, K., Moorthi, P. S., & Kumar, R. S. (2017). A study of image processing in agriculture. Int. J. Advanced Networking and Applications, 09(01), 5. Retrieved from https://www.ijana.in/papers/V9I1-4.pdf

Pujari, D., Yakkundimath, R., & Byadgi, A. S. (2015). Image Processing Based Detection of Fungal Diseases in Plants. *Procedia Computer Science*, *46*, 1802–1808. doi:10.1016/j.procs.2015.02.137

Puno, John Carlo & Sybingco, Edwin & Dadios, Elmer & Valenzuela, Ira & Cuello, Joel. (2017). Determination of soil nutrients and pH level using image processing and artificial neural network. 1-6. . doi:10.1109/HNICEM.2017.8269472

Raichaudhuri, Rittika & Sharma, Rashmi. (2017). On Analysis of Wheat Leaf Infection by Using Image Processing. 468. 569-577. . doi:10.1007/978-981-10-1675-2_56

Ramos, P. J., Prieto, F. A., Montoya, E. C., & Oliveros, C. E. (2017). Automatic fruit count on coffee branches using computer vision. Computers and Electronics in Agriculture, 137, 9e22.

Rau, A. J., Sankar, J., Mohan, A. R., Krishna, D. D., & Mathew, J. (2017). IoT based smart irrigation system and nutrient detection with disease analysis. 2017 IEEE Region 10 Symposium (TENSYMP), 1-4.

Roopaei, Mehdi & Rad, Paul & Raymond Choo, Kim-Kwang. (2017). Cloud of Things in Smart Agriculture: Intelligent Irrigation Monitoring by Thermal Imaging. IEEE Cloud Computing. 4. 10-15. . doi:10.1109/MCC.2017.5

S. ANITHA &, Dr.V.RADHA. (2010). Comparison of Image Preprocessing Techniques for Textile Texture Images. International Journal of Engineering Science and Technology. 2.

Sara Ghods and Vahhab Shojaeddini, "A novel automated image analysis method for counting the Population of whiteflies on leaves of crops'', J. Crop Prot., 2016.

Schoeman, L., Williams, P., Plessis, A., & Manley, M. (2016). X-ray micro-computed tomography (μCT) for non-destructive characterisation of food microstructure. *Trends in Food Science & Technology*, *47*(1), 10–24. doi:10.1016/j.tifs.2015.10.016

Sereshti, H., Poursorkh, Z., Aliakbarzadeh, G., Zarre, S., & Ataolahi, S. (2017). An image analysis of TLC patterns for quality control of saffron based on soil salinity effect: A strategy for data (pre)-processing. Food Chemistry. doi: PubMed doi:10.1016/j.foodchem.2017.07.012

Sharma, K. (2018). *Hement & Kumar, Shiv* (pp. 885–890). Soil Classification & Characterization Using Image Processing; doi:10.1109/ICCMC.2018.8488103

Shit, P., Bhunia, G., & Maiti, R. (2015). Soil crack morphology analysis using image processing techniques. *Modeling Earth Systems and Environment*, *1*(4), 1–7. doi:10.1007/s40808-015-0036-z

Silwal, A., Gongal, A., & Karkee, M. (2014). Identification of red apples in field environment with over the row machine vision system. *Agricultural Engineering International: CIGR Journal*, *16*(4), 66–75.

Sonka, M., Hlavac, V., & Boyle, R. (1993). Image pre-processing. In *Image Processing, Analysis and Machine Vision*. Boston, MA: Springer; doi:10.1007/978-1-4899-3216-7_4.

Srisha, Ravi & Khan, Am. (2013). Morphological Operations for Image Processing: Understanding and its Applications.

Srunitha, K & Sivasamy, Padmavathi. (2016). Performance of SVM classifier for image based soil classification. 411-415. . doi:10.1109/SCOPES.2016.7955863

Stein, M., Bargoti, S., & Underwood, J. (1915). 2016. Image based mango fruit detection, localisation and yield estimation using multiple view geometry. *Sensors (Basel)*, *16*, 1–25.

Studio encoding parameters of digital television for standard 4:3 and wide screen 16:9 aspect ratios, International Telecommunication Union Recommendation BT.601-7, 2011.

Sun, Y., Cheng, H., Cheng, Q., Zhou, H., Li, M., Fan, Y., ... Lammers, S. (2017). A smart-vision algorithm for counting whiteflies and thrips on sticky traps using two-dimensional Fourier transform spectrum. *Biosystems Engineering*, *153*, 82–88. doi:10.1016/j.biosystemseng.2016.11.001

Vibhute, Anup &, S.K.Bodhe. 2012. Application of Image Processing in Agriculture:Survey, A. (•••).. . *International Journal of Computers and Applications*, *52*, 34–40. doi:10.5120/8176-1495

Sushma R. Huddar and SwarnaGowri, Keerthana K., Vasanthi S and Sudhir Rao Rupanagudi, "Novel Algorithm for Segmentation and Automatic Identification of Pests on Plants using Image Processing'', ICCCNT'12, July 2012.

Tajane, V., & Janwe, N. J. (2014). Medicinal Plants Disease Identification Using Canny Edge Detection Algorithm, Analysis and CBIR. *International Journal of Advance Research in Computer Science and Software Engineering*, *4*(6), 530–536.

Tsai, C.-M., & Yeh, Z.-M. (2008). Contrast enhancement by automatic and parameter-free piecewise linear transformation for color images. Consumer Electronics. *IEEE Transactions on.*, *54*, 213–219. doi:10.1109/TCE.2008.4560077

Valliammai, N., & Geethaiakshmi, S. N. "Multiple noise reduction using hybrid method for leaf recognition," 2012 International Conference on Devices, Circuits and Systems (ICDCS), 15-16 March 2012, Coimbatore, India. doi:10.1109/ICDCSyst.2012.6188689

Vibhute, A., & Bodhe, S. K. (2012). Article: Applications of Image Processing in Agriculture: A Survey. *International Journal of Computers and Applications*, *52*, 34–40. doi:10.5120/8176-1495

Wang, Q.I. & Nuske, Stephen & Bergerman, Marcel & Singh, Sanjiv. (2012). Automated Crop Yield Estimation for Apple Orchards. Experimental Robotics: The 13th International Symposium on Experimental Robotics. 88. . doi:10.1007/978-3-319-00065-7_50

Chunhua Zhu, Jiaojiao Wang, Hao Liu, and Huan Mi, "Insect Identification and Counting in Stored Grain: Image Processing Approach and Application Embedded in Smartphones," Mobile Information Systems, vol. 2018, Article ID 5491706, 5 pages, 2018. . doi:10.1155/2018/5491706

Zulkifli Bin Husin., Abdul Hallis Bin Abdul Aziz, Ali Yeon Bin Md Shakaff Rohani Binti S Mohamed Farook, "Feasibility Study on Plant Chili Disease Detection Using Image Processing Techniques", 2012 Third International Conference on Intelligent Systems Modelling and Simulation.

Chapter 3
Expert System Design for Diagnosis of Diseases for Paddy Crop

Sreekantha Desai Karanam
NMAM Institute of Technology, India

Deepthi M. B.
NMAM Institute of Technology, India

ABSTRACT

India has the second largest area of arable (agricultural) land on this earth with heterogeneous agroclimatic regions across the country. India has the potential to grow a wide range of agricultural crops and varied raw material base for food processing industry. The paddy crop yield/hector of land is highest in Egypt is 9.5, while India is producing only 2.9. India's lower paddy crop productivity/hector and higher cost of production is a major concern for farmers. There are various reasons for India's low paddy crop yield, such as lack of mechanization, not adopting to modern method of farming, small land holdings, poor pests, and disease management. The recent survey discovered that there is huge gap in demand and supply in crop production and is likely to hit more than 15% by 2020, with the gap worsening to 20-25% by 2025. Researchers aimed to address this low crop yield issue by designing an expert system. This expert system helps the farmers by identifying and predicting the diseases for paddy crop to enhance crop yield and to reduce the supply and demand gap.

INTRODUCTION

India is an agriculturally based country and the agriculture is one of the most important economic sectors. About 75% Indian population is living in rural areas and depend on agriculture and allied activities. Food and Agriculture Organization (FAO) reported that 80% of the world paddy production comes from only seven countries. China contributes to 32.7% and India's share is 26.0% to world paddy production. The

DOI: 10.4018/978-1-5225-9632-5.ch003

cost of producing paddy crop per quintal in India also varies from Rupees 639 in Andhra Pradesh, Rupees 783 in Madhya Pradesh and Rupees 435 in Bihar. These production and cost statistics highlights the urgent need to enhance the paddy crop yield/ha and to optimize the cost of production. There are several reasons for having low productivity in India, which is dominated by small farmers, having less than two acres of land, sometimes in hill stations. The present paddy crop cultivation machines are designed for large farmers in plane lands. Farmers are not using modern tools and methods of cultivation. The modern agriculture requires integration of information and knowledge from various sources. Nowadays the farmers usually depend on agricultural scientists and experts to get better information for their decision making. Agricultural scientists are not always available in remote areas when and where the farmers need them. To overcome these problems, the expert systems are developed as a powerful tool in agriculture to assist the farmers. The farmer can avail the services of these expert system using smart phones.

An expert system is an application software which imitates the thought process of human expert to achieve the performance equivalent to human expert in a particular problem domain. Expert systems is a subset of Artificial Intelligence domain. Authors have carried out an extensive survey of literature and identified several existing expert systems helping the farmers in taking very important decisions for diagnosis of diseases, disorders, pests, crop selection and crop management. The researchers have designed an expert system prototype for the diagnosis of various diseases affecting paddy crop by observing their disease symptoms. The step by step procedure adopted in designing rule based expert system is described in this chapter. This paddy crop disease expert system has three building blocks i.e Inference engine, Knowledge base, and User-interface. A user interface (UI) component is used enables communication between the system and the users. It allows the users to query the system or answering questions and to receive system advices. The collection of facts about the paddy crop diseases are designed in the form of rules and stored in rule base. The tacit knowledge is derived through interactions with domain human experts and later this knowledge is encoded in to a knowledge base. Users enter their inputs and questions through user interface. These questions are interpreted and processed by inference engine using the knowledge from knowledge base and rules. This expert systems provide advice to farmers based on user's query and symptoms of the crop entered.

REVIEW OF LITERATURE

The precise identification of paddy crop diseases and pests is very essential for enhancing the rice crop yield and quality cultivation (Peng et al., 2010). The process of diagnosis of paddy crop diseases is very complex. Authors have applied Back Propagation Algorithm (BPA) of neural network technology to design an expert system for diagnosis of diseases, which is more efficient in processing incomplete and vague information. Authors (Li, Zhang & Yang, 2012) have proposed the characteristics of hybrid approach, such as self-learning ability of neutral network, inference ability of fuzzy systems and knowledgebase of expert system for prediction of crop growth. Authors have analysed growth rate of crop and the performance of expert systems using simulation and testing. The results obtained demonstrated that this practice is effective in the high crop yield management and better quality. Authors (Kaur, Singh Rekhi & Nayyar, 2013) have discovered that currently majority of Fortune 1000 companies are developing expert systems for enhancing the quality, efficiency and competitive leverage in their day to day operations. The expert systems are being used in scientific, business and industrial applications such as, to discover oil or mineral deposits, control various space crafts and diagnosing medical diseases.

Researchers (Das & Nayak, 2013) have discussed the application of expert systems in agriculture as a tool to provide information to the farmers based on their needs using the knowledge captured from domain experts. This expert system provides the farmers with context sensitive online information based on their needs and questions. Researchers (Kannan & Hemalatha, 2012) have designed Agro Genius: A nascent expert system designed for farmers to ask questions and get clarifications, which is implemented by applying data mining techniques. This Agro Genius expert system provides information that enables the farmers to grow crops having high yield and create awareness on organic farming methods. This system helps farmers to query and get clarifications associated to agriculture for better yield before cultivation. Milind K. Tatte and Mangesh K. Nichat (2013) have developed expert system model for diagnosis of diseases in the rice plant. The knowledge base contains the knowledge about the different diseases of rice plant represented in separate sections. The system architecture presented here is an integrated system with interactive user interface, control and coordinating units, expert system shells, and structured knowledge representations. Modern agricultural production process (Sriram & Philip, 2014) is demanding information derived from many different sources. Today's farmers depend on agricultural experts and scientists to get the information for decision making. Unfortunately, the timely help agricultural experts will not be always available to farmers when they need them especially in remote areas. The online web-based expert systems deployed are powerful tools in agriculture that are easy to access anywhere, any time of the day by the farmers using smart phones. Authors have also explored various experts systems implemented from the year 1984 to 2014.

The authors (Deepthi & Sreekantha, 2017) have carried out the survey of literature on existing expert systems related to various crops in agriculture from the year 1985 to 2016. This survey paper has been presented in IEEE international conference. These expert systems are very useful for the farmers to get an advice about the application of fertilizers, pest management and the management of the crop yield.

Some of the existing expert systems studied are

1. AgPest is used for diagnosis of rice and wheat crops
2. Web Based Expert System for leaf diseases of cereals
3. LIMEX for diagnosis of lime crop management
4. NEPER for wheat cultivation
5. CUPTEX for cucumber cultivation

An agricultural expert system has been developed (Pradeep Kumar Singh, April 2018) by integrating the accrued expertise of many disciplines such as entomology, horticulture, soil management into a framework that to solve specific problems and needs of farmers. This system helps in making the best decisions for raising a successful crop. Tucha Kedir Elemo, Meeta Kumar (May 26, 2018) have developed Intelligent Expert System for diagnosis of disease in wheat plant. The proposed system uses Jess (Java Expert System Shell) for diagnosis and treatment of wheat plants. The Expert System for Rice Plant Disease Diagnosis, ESforRPD2 was developed (Agus et al., 2019). This system is based on the pest and disease experiences of the rice experts. This Expert System can detect 48 symptoms and 8 types of diseases of rice plants from 16 data tests with a sensitivity of 87.5%., Currently, ESforRPD2 has only been tested with data from the Samarinda region of Indonesia.

PROBLEM SUMMARY

Researchers have studied diseases affecting the paddy crop and gathered the information of about 43 paddy crop diseases with their symptoms. These symptoms are then classified into different groups based on infected parts, types of defects, color and shapes. These symptoms are treated as input parameters for disease prediction. The precise paddy crop disease identification is the output of expert system. The paddy crop disease stages are classified as preliminary, matured, serious and dead stage based these symptoms. Authors have discovered that some paddy crop diseases are having some common symptoms. Researchers have combined four disease symptoms with similar or same symptoms. Authors have visited Zonal Agricultural Research Centre in Bramhavar town of Udupi District in Karnataka to study the paddy crop diseases and to evaluate their expert system prototype.

METHODOLOGY

Researchers have followed the expert system approach to study the diseases and designed an expert system prototype using following steps.

Step by step procedure for designing of the expert system prototype

1. Studying the symptoms of diseases of paddy crop by literature survey, interacting with paddy crop scientists, farmers.
2. Identifying the types of diseases, their symptoms and to encode this knowledge into knowledge base
3. Classifying the symptoms into different types of stages of diseases such as preliminary, matured, serious and dead stage
4. Designing the fuzzy rules and to build rule base from the knowledge acquired through interactions with human domain experts
5. Building the expert system prototype, training and testing using Fuzzy logic toolbox and Neural network toolbox of Matlab software

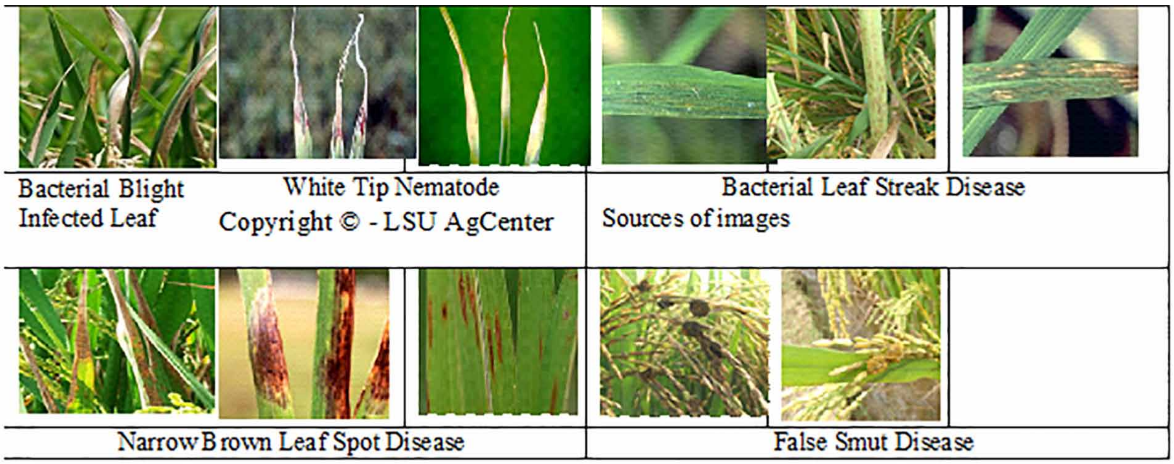
Step 1: Study of Paddy Crop Diseases and Symptoms

Researchers have carried out an exhaustive survey of applications of expert system to various crops in agriculture and diseases by studying the literature and discussion with experts. Authors have studied the literature on agriculture from the year 1985 to 2016. This study enabled authors to publish a research paper in IEEE conference (Sreekantha, Deepthi & Puneeth, 2017).

Step 2: Identifying Types of Diseases, Their Symptoms

Researchers have gathered the information of 43 paddy crop diseases and their symptoms. The symptoms are classified into different stages. Infected parts, types of defects, color and shapes are the input variables. Diseases and various stages are the output variables. Figure 1 shows the images of typical sample paddy crop diseases with symptoms of infected parts

Figure 1. Paddy crop disease symptoms and diseases
Source: http:/irri.org/decision-tools/rice-doctor/rice-doctor-fact-sheets/item/ bacterial-blight



Step 3: Classifying the Symptoms Into Different Types of Stages of Diseases

The symptoms are classified into different groups based on infected part, types of defects, color and shapes. The paddy crop disease symptoms and diseases stages are organized as depicted in Table 1

Step 4: Rule Base Design

Researchers have organized the knowledge acquired from the literature study, interactions with experts and field study of 43 diseases in to about 150 rules. Sample 30 rules are shown into Table 2.

Step 5: Designing an Expert System Prototype for Diseases Diagnosis Using Fuzzy Logic Toolbox

Authors have implemented an expert system prototype using MATLAB software toolbox, denotes MATrix LABoratory is a latest mathematical software package. It has graphical, numerical and programming capabilities. It has many built in functions to do numerous operations, and there are toolboxes that can be added to enhance its utility.

Fuzzy Logic Toolbox

The Fuzzy Logic toolbox has a group of functions based on MATLAB numerical computing environment. It gives GUI tools to create and edit fuzzy inference systems within the structure of MATLAB.

This toolbox depends heavily on GUI tools to enables to achieve the work, in spite of that one can work on totally from the command line also. The fuzzy logic toolbox contains the tools such as FIS editor, Membership function editor, Rule editor, Rule viewer and Surface viewer. The membership functions that are associated with each symptom variable are defined by the Membership Function Editor. The list of rules that represent the system performance are entered into the system using Rule Editor. The

Table 1. Paddy crop disease symptoms and diseases stages

Sl. No	Disease name	Preliminary stage	Matured stage	Serious stage	Dead stage
1	White Tip.	Leaf tip becoming crinkled and twisted.	Diseased plants produce small panicles, lacks vigor and stunted.	Affected panicles show distorted small kernels, distorted glumes, and high sterility.	Affected grains become chaffy.
2	Bacterial Blight.	The leaf tip water soaked lesions move downwards on the leaf edge.	Infected leaves roll-up and the color turned into grayish green.	Leaves are wilting and the color turned into yellow to straw-colored.	Whole seedlings dies and dry up.
3	Leaf Scald.	Water-soaked appearance and gray-green color.	Lesions develop a dark reddish brown and light color chevron pattern.	Lesion edges of the lesion become yellow to gold.	An affected leaf has scalded appearance and leaves become dry and color turned into straw colored
4.	Bacterial Leaf Streak.	Symptoms on the leaf blades are water soaked, dark greenish, narrow, and various lengths of interveinal streaks.	Color of the lesion turn yellow-orange to brown and enlarge. On the lesion bacterial exudates of amber color is seen.	Due to numerous streaks the large areas of leaf become dry.	Lesion color turned into brown to grayish white and leaves become brown and dry.
5.	Narrow Leaf Spot.	Lesions are narrow; elliptical to linear, short, and brown color usually occurs on leaf blades.	Lesion becomes shorter, dark brown, and narrower on resistant varieties.	Lesions become lighter brown with gray-necrotic center and wider on susceptible varieties.	On the lesion abundant sporulation is occurred.
6.	False Smut.	Grains discoloration.	Grains are transformed into mass of yellow fruiting bodies.	Smut balls in greenish black color with velvety appearance.	Smut Balls Bursts and becomes Black in Color.
7.	Rice Grassy Stunt.	Plants are stunted and leaf color is yellowish green that are narrower and shorter than normal.	A leaf has a numerous patches or a small rusty spot which is then merge into blotches.	Mottled appearance on leaves.	Plants that fail to produce panicles.
8.	Brown spot.	Seedling has brown and circular lesion.	Roots are black in color.	On older plants the lesion has reddish brown margin, oval/ circular, with grayish center.	Black/brown spots on grain and large areas of leaves are dry.
9.		Lesions on leaf sheath of lower leaves near the water line.	Lesions develop below the leaf collar as water-soaked spots, green-gray, oval to elliptical.	Lesions become grayish white to light (tan) with brownish borders and dry.	Initially Sclerotic appear on or near the lesions, white in color and turn into dark brown.
10.	Sheath Rot.	Irregular lesions/spot, with gray center and dark reddish-brown margin.	Unmerged florets and panicles rot color turned into red-brown to dark-brown.	Inside the affected young panicles and sheaths the whitish powdery growth is appeared.	Infected panicles shriveled, sterile or grain is partially filled.

Surface Viewer and the Rule Viewer are read-only tools. Authors selected the symptoms of the plant's infected part as input parameters and used membership functions for each and every input symptom. The predicted disease as output parameter. The leaves and seedlings are the members of infected part variable. The current variable field displays the name, type and range of the variable. The authors have designed about 150 rules for paddy crop diseases and some ten rules areas illustrated in Figure 2.

Table 2. Rules and conclusions for different types of diseases in paddy crop

Sl. No	Rules	Conclusions
1.	If a crop is rice and tip of the leaf is twisted.	Preliminary stage symptom for the disease White Tip.
2.	If a crop is rice and tip of the leaf is crinkled.	Preliminary stage for disease White Tip.
3.	If a crop is rice and plants are stunted and produce small panicles.	Matured stage symptom for disease White Tip.
4.	If a crop is rice and high sterility in affected panicles and small and distorted kernels and glumes.	Serious stage symptom for disease White Tip.
5.	If a crop is rice and the affected grains become chaffy.	Dead stage symptom for disease White Tip.
6.	If a crop is rice and water-soaked lesions move from tip downwards on the leaf edge	Dead stage symptom for disease White Tip
7.	If a crop is rice and color of leaves turn into grayish green and roll up.	Matured stage symptom for disease Bacterial Blight.
8.	If a crop is rice and the leaf is wilt and turned into yellow to straw coloured.	Serious stage symptom for disease Bacterial Blight.
9.	If a crop is rice and the entire seedlings to dry up and dies.	Dead stage symptom for disease Bacterial Blight.
10.	If a crop is rice and color is gray green and water soaked appearance	Preliminary stage symptom for disease Leaf Scald.
11.	If a crop is rice and lesions developed as a dark reddish brown and light color chevron pattern.	Matured stage symptom for disease Leaf Scald.
12.	If a crop is rice and the color of leaf edge is yellow to gold.	Serious stage symptom for disease Leaf Scald.
13.	If a crop is rice and infected leaves turn straw-coloured and dry-up and appear scalded	Dead stage symptom for disease Leaf Scald.
14.	If a crop is rice and shape is narrow and color is dark greenish and water-soaked.	Preliminary stage symptom for disease Bacterial Leaf Streak.
15.	If a crop is rice and shape is narrow and color is dark greenish and various lengths of interveinal streaks on leaf blade and water soaked.	Preliminary stage symptom for disease Bacterial Leaf Streak.
16	If a crop is rice and oval or circular lesion with gray center and older plant has reddish-brown margin.	Serious stage symptom for disease Brown Spot on seedling.
17	If a crop is rice and lesions are dry and brownish borders and become grayish white to tan.	Serious stage symptom for disease Sheath Blight.
18	If a crop is rice and several tiny black and white mycelium and sclerotic visible inside the infected stem	Matured stage symptom for disease Stem Rot.
19	If a crop is rice and delayed flowering and shape of the panicle is small and not completely exerted.	Serious stage symptom for disease Rice Tungro Virus.
20	If a crop is rice and oblong to irregular and dark green and water-soaked lesion and turns to gray brown or brown.	Matured stage symptom for disease Bacterial Sheath Brown Rot.
21	If a crop is rice and part is node and color is blackish to grayish-brown and defect is lesion.	Dead stage symptom for disease Bacterial Sheath Brown Rot.
20	If a crop is rice and oblong to irregular and dark green and water-soaked lesion and turns to gray brown or brown.	Matured stage symptom for disease Bacterial Sheath Brown Rot.
21	If a crop is rice and part is node and color is blackish to grayish-brown and defect is lesion.	Dead stage symptom for disease Bacterial Sheath Brown Rot.
22	If a crop is rice and elliptical or spindle shaped lesions on leaves and necrotic or red to brownish border and whitish to gray-center.	Serious stage symptom for disease Blast (leaf and collar).

continues on following page

Table 2. Continued

Sl. No	Rules	Conclusions
23	If a crop is rice and crown area is decayed with soft rotting and dark brown or black color and discolored streaks.	Dead stage symptom for disease Blast (leaf and collar).
24	If a crop is rice and older plant have purple brown blotches on leaves and yellow to bronze lower leaves floating on surface.	Serious stage symptom for the disease Bronzing.
25	If a crop is rice and elongated lesions on leaves near the tip and margin of the leaf and appear as water soaked.	Preliminary stage symptom for the disease Bacterial Leaf Blight.
26	If a crop is rice and lesion is oval or round and pale tan or white spots and narrow red or reddish brown margin.	Preliminary stage symptom for disease Stackburn.
27	If a crop is rice and lesions appear as a water-soaked spots on the rice plant near the water line.	Preliminary stage symptom for disease Aggregate Sheath Spot.
28	If a crop is rice and leaf blade white to dark brown or pale yellow veins are developed and sheaths galls on outer surface of leaf sheath and underside of leaf blades.	Serious stage symptom for the disease Rice Ragged Stunt.
29	If a crop is rice and panicles upright and slightly bent over because of sterility and not falling over.	Preliminary stage symptom for disease Straight head.
30	If a crop is rice and green stems and discoloration of grains and infected grains unevenly spreads on the panicle.	Serious stage symptom for disease Bacterial Panicle Blight.

Authors have selected infected parts as an input variables in Fuzzy Logic Designer and Membership Function Editor tools as shown in Figure 3. The infected parts of the crop leaves and seedlings are the members of infected part variable. The current variable field displays the name, type and range of the variable in Membership Function Editor tool. Click on a line to choose it and one can modify any of its characteristics such as type, numerical parameter and name. In current variable field, the type field has a pop-up menu and it is used to modify the type of the current membership function.

Rule Editor

The GUI rule editor tool is useful for entering rules. The selection menus helps to select the symptoms of one particular disease and it display rules in the form of IF….AND….THEN statement. Rule editor tool is used to create and edit the rules by using the GUI buttons like delete rule, add rule and change rule. It has two connections like OR and AND. Rule editor enables us to construct rules automatically by selecting and clicking one item from each input variable box, one item from each output variable box and one connection item .

For example: “IF Infected Parts are Leaves AND Symptom is RollUp AND Color is Grayish Green THEN Stage is Matured and Disease is Bacterial Blight.”

Output Design: Disease Stage Prediction

Membership Function Editor is used to identify the infected part disease and its stage accurately as output as shown in Figure 4. The disease stage is selected as Preliminary, Matured, Serious and Dead which are the four stages of the bacterial blight disease as a membership functions.

Figure 2. Fuzzy logic designer and membership functions for paddy crop diseases symptoms

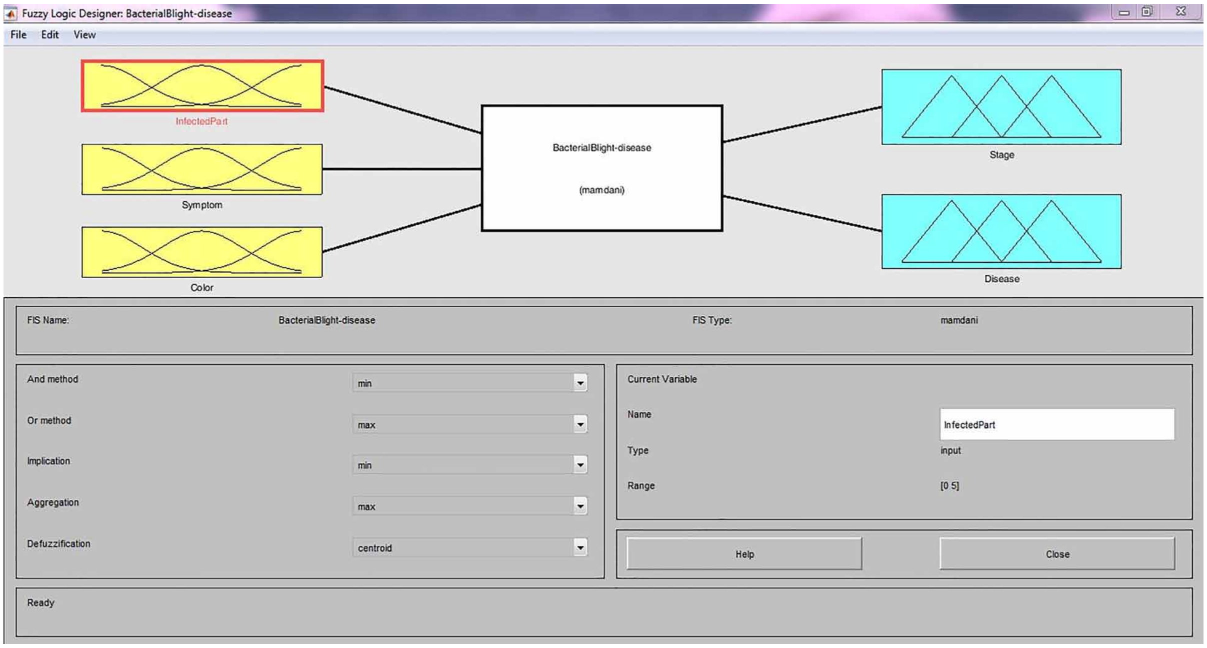


Figure 3. Rule Editor showing the rule designed for crop disease prediction

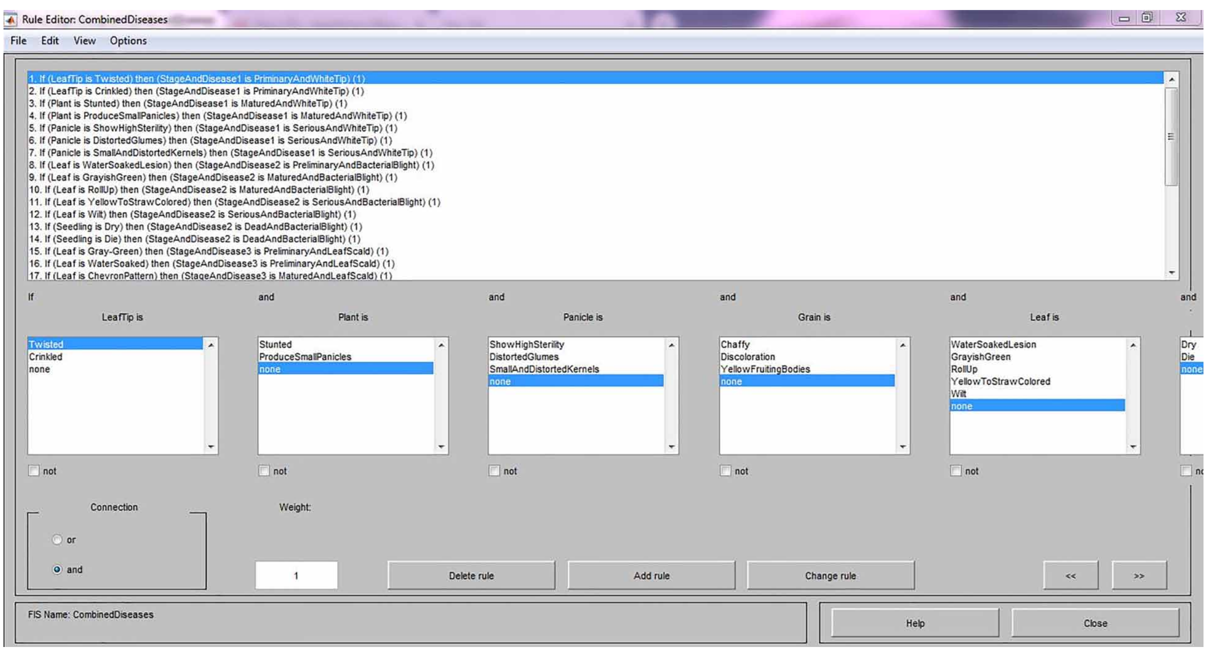
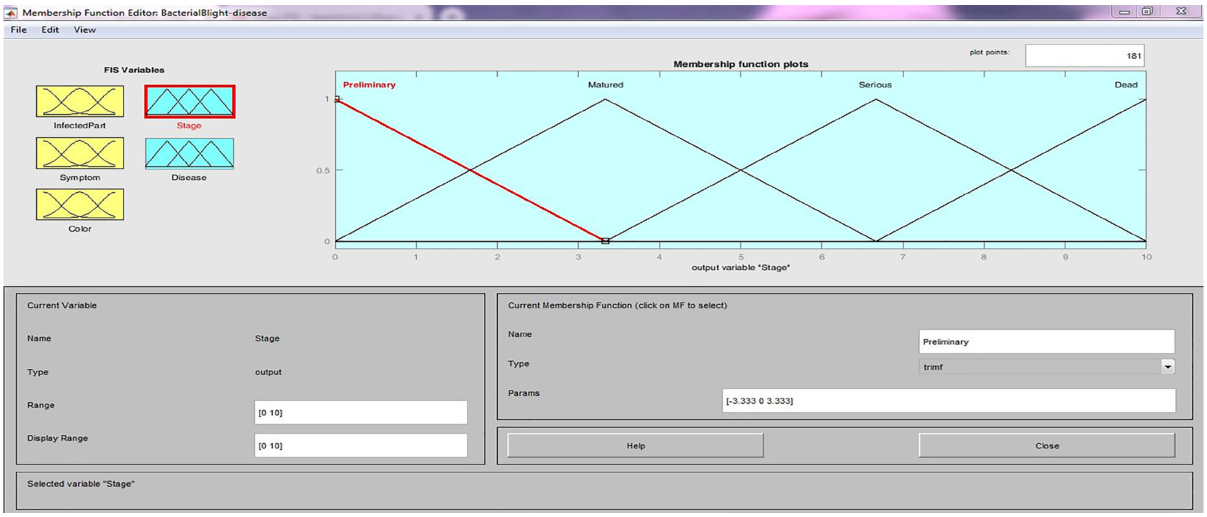
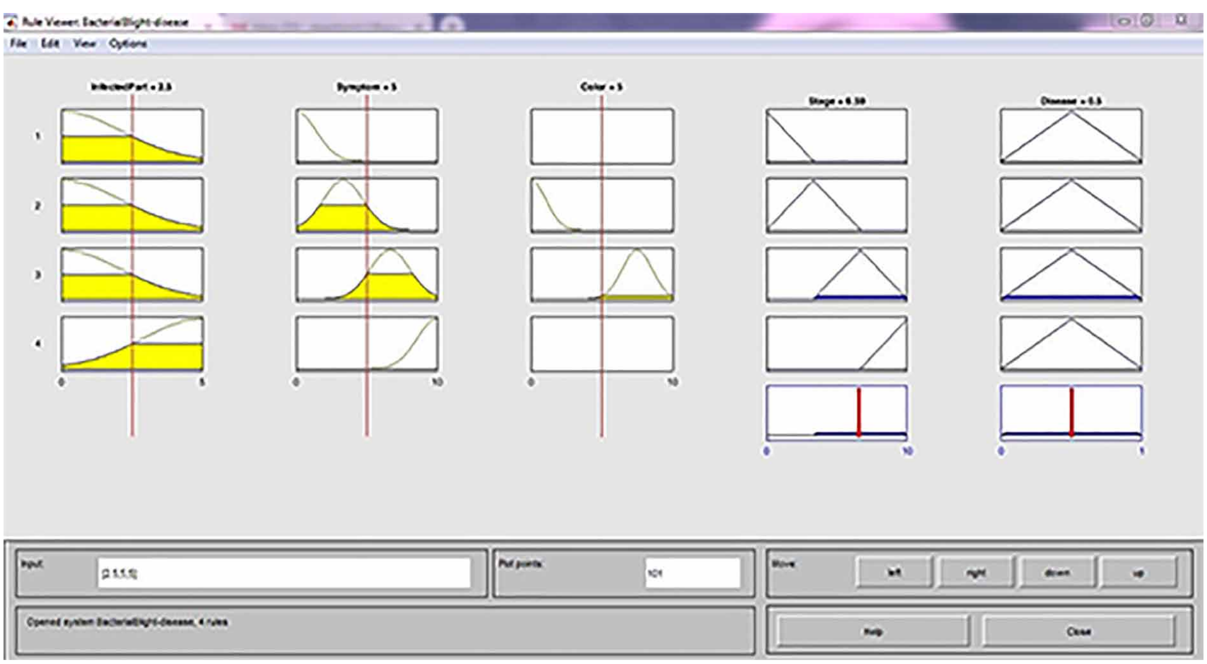


Figure 4. Membership function editor for different stages of bacterial blight disease



Rule Viewer The rule viewer tool of Fuzzy logic toolbox as shown in Figure 5 depicts how the output variable changes when the symptoms of disease increase from preliminary stage to dead stage. Each row of plots represents one rule. Each column of plots (yellow) displays how the input variables are utilized as a part of the rules and each column of plots (blue) displays how the output variables are utilized as a part of the rules. Slide the red vertical line above the input values and it generates new output values.

Figure 5. Rule viewer for bacterial blight disease



Surface Viewer

Surface Viewer tool has X, Y, and Z-axis in the Figure 6. One can select Infected Part as X-axis, Color as Y-axis and Disease stage as Z-axis. If the user changes the X, Y, and Z-axis then the shape of the surface viewer changes. This plot demonstrates the output surface for any one or two inputs to the system versus any output of the system. The pop-up menu is used to set the input and output variables. This tool is used to visualize the relationship between input and output variables.

Matlab Neural Network toolbox implementation

Major Steps in Artificial Neural Network (ANN) Design are

The steps followed in implementation of ANN using Artificial Neural Network Design are shown Figure 7

1. Inputs data about the crop has to be collected
2. The data is converted into decimal numbers randomly generated
3. Normalize the input to check the uniformity of the data.
4. The normalized input data is divided in to two subsets, training data is 85% and testing data 15% to be given to the neural network.
5. Artificial Neural Network (ANN) is trained using training data set. ANN performance is tested later using testing data
6. The predicted output is recorded and mean square error is computed.

The prediction accuracy is calculated for given data. The random decimal numbers are generated for the input data and output, because it is easy to analyze the decimal number data by the neural networks. Normalization of the input data is carried out to ensure uniformity between the data values. Normalization is carried out to have same ranges of value for each of the input using the following formula the data is shown in Figure 8.

Figure 6. Surface viewer for bacterial blight disease

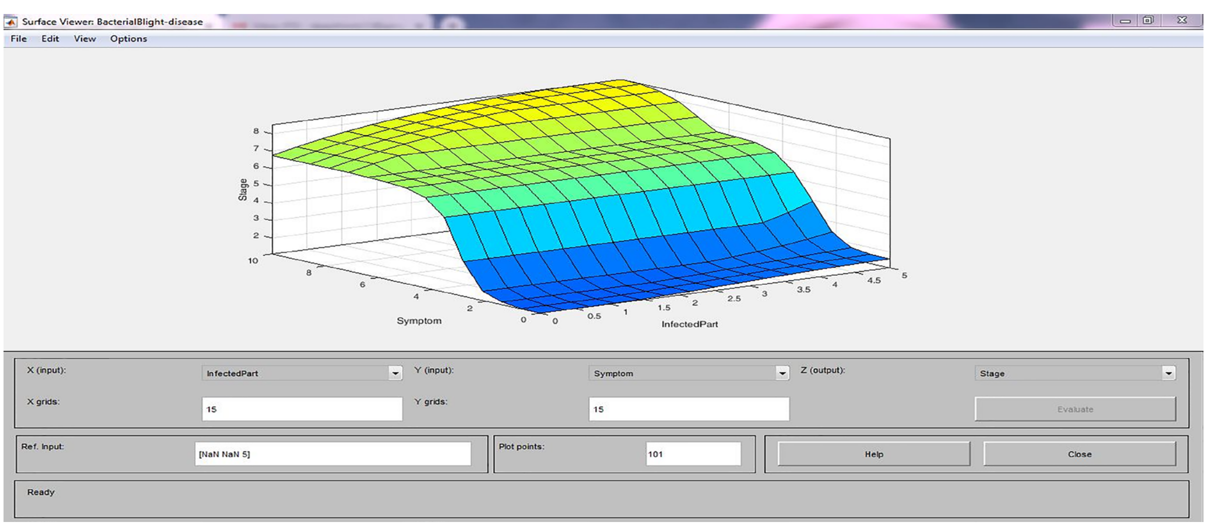
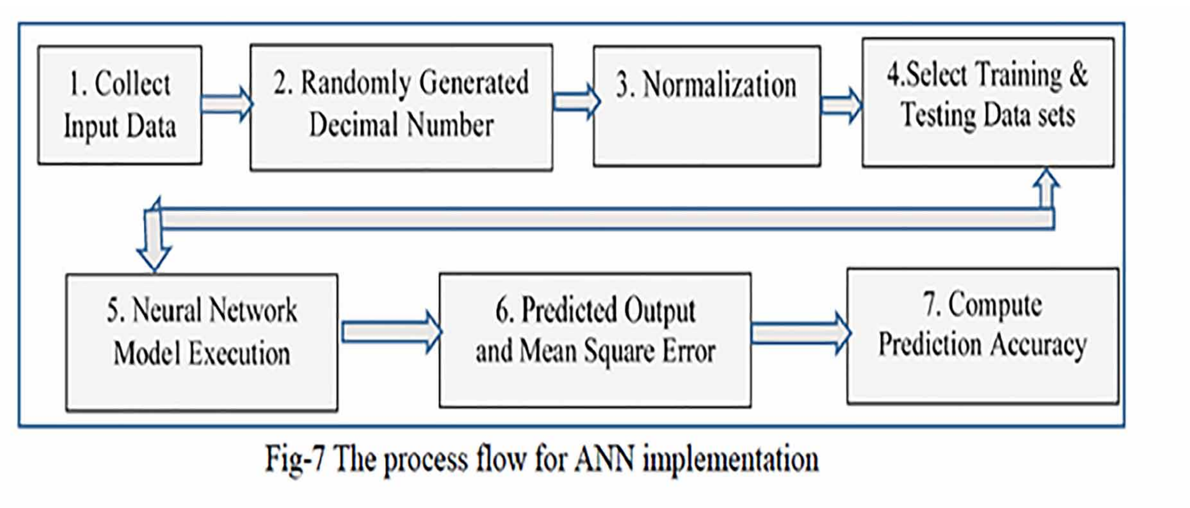


Figure 7. The process flow for ANN implementation



Normalization = number / maximum number in that column

The dataset should be divided into training and testing data sets. For large data sets the data is divided by 85:15 ratios. Here, out of 46 data 38 data for training and remaining 7 data for testing shown in Figure 9.

Formula for Accuracy = (Number of Correct Samples/Number of Total Samples)*100

Figure 8. Input data sets before normalization

	A	B	C	D	E	F	G
1	infected part	color	shape	type of defects		disease	stage
2	1			1		1	1
3	2			2		1	4
4	3			3		2	2
5	3			4		2	3
6	3			5		3	4
7	3	1		6		4	1
8	3			7		5	1
9	3			8		5	4
10	2		2	9		6	2
11	2	7		10		6	3
12	3			11		7	2
13	3			12		7	3
14	2	2		7		8	4
15	4	3		13		9	2
16	5	4		14		10	2
17	3	5	3			11	1
18	2			15		12	4
19	3	6		16		13	1
20	6			17		14	1
21	9	11		18		15	1
22	8			19		16	3
23	3			20		17	1
24				21		18	3
25				22		19	1
26	3	9		23		20	3
27				24		21	2
28	7			25		22	1
29	12			26		22	2
30	11	11	5	16		23	1
31		11	2	27		24	3
32			1	28		25	1
33	4			29		26	2
34	9	14		30		27	2
35			4	31		28	2
36				32		28	4
37	8	10		33		29	1
38	8		6	34		30	1
39		15		35		31	1
40				36		32	1
41				37		32	4
42				38		33	2
43	5	16				34	1
44		12		39		35	3
45	2			40		36	1
46	3	17				36	2

Figure 9. Normalized input data set for training

	A	B	C	D	E	F	G
1	0.0833	0	0	0.025		0.0277	0.25
2	0.1666	0	0	0.05		0.0277	1
3	0.25	0	0	0.075		0.0555	0.5
4	0.25	0	0	0.1		0.0555	0.75
5	0.25	0	0	0.125		0.0833	1
6	0.25	0.0588	0	0.15		0.1111	0.25
7	0.25	0	0	0.175		0.1388	0.25
8	0.25	0	0	0.2		0.1388	1
9	0.1666	0	0.3333	0.225		0.1666	0.5
10	0.1666	0.4117	0	0.25		0.1666	0.75
11	0.25	0	0	0.275		0.1944	0.5
12	0.25	0	0	0.3		0.1944	0.75
13	0.1666	0.1176	0	0.175		0.2222	1
14	0.3333	0.1764	0	0.325		0.25	0.5
15	0.4166	0.2352	0	0.35		0.2777	0.5
16	0.25	0.2941	0.5	0		0.3055	0.25
17	0.1666	0	0	0.375		0.3333	4
18	0.25	0.3529	0	0.4		0.3611	0.25
19	0.5	0	0	0.425		0.3888	0.25
20	0.75	0.647	0	0.45		0.4166	0.25
21	8	0	0	0.475		0.4444	0.75
22	0.25	0	0	0.5		0.4722	0.25
23	0	0	0	0.525		0.5	0.75
24	0	0	0	0.55		0.5277	0.25

Artificial Neural Network Experiments

Artificial neural network toolbox in MATLAB facilitates reading the input dataset, and setting the number of hidden layer neurons, number of epochs, goal, transfer functions and type of training algorithms. The neural network experiments with 5 to 25 hidden neurons are carried out. The predicted output and (Mean Square Error) MSE for two outputs using tansig and logsig transfer functions and four types of training functions namely trainlm, trainbfg, trainsig and traincgf is carried out. Finally, the prediction accuracy for training and testing datasets are computed. The error tolerance is set to 0.05 value and if the output value is above 0.05, it is neglected and if it is less than or equal to 0.05 then it should be counted for calculating total accuracy. Accuracy is calculated by using formula. The two sigmoid transfer functions are used to train the network such as logsig and tansig. The log sigmoid transfer functions (logsig) and hyperbolic tangent sigmoid transfer function (tansig) are used to calculate a layers output from its net input. It trains the network by setting 5 to 25 number of hidden neurons, epochs which are set to be 1000, error and prediction accuracy for every normalized input. In the Figure 10 the Best value of accuracy is considered is better for training and testing by using tansig function is 86.8421% for training and 71.4286% for testing. The LogSig function accuracy is 81.5789% for training and 71.4286% for testing using logsig and trainbfg.

Figure 11 shows the prediction accuracy of the output disease by comparing different types of training algorithm. The accuracy is calculated by considering highest prediction accuracy with the lowest hidden neurons. The accuracy is considered better for training and testing by using tansig and traincgf with the 100% for training and 100% for testing. Accuracy is 100% for training and 100% for testing using logsig and trainlm.

Figure 10. Performance analysis of different algorithms during training and testing of data

Training o1				
	trainlm	*trainbfg*	*trainscg*	*traincgf*
5	100	100	100	**100**
6	**100**	100	100	100
8	100	100	**100**	100
18	100	**100**	100	100

Testing o1				
	trainlm	*trainbfg*	*trainscg*	*traincgf*
5	71.4286	71.4286	71.4286	**71.4286**
6	**100**	85.7143	71.4286	57.1429
8	100	71.4286	**100**	57.1429
18	71.4286	**100**	71.4286	71.4286

Training o2				
	trainlm	*trainbfg*	*trainscg*	*traincgf*
5	**89.4737**	**81.5789**	92.1053	76.3158
6	92.1053	78.9474	71.0526	**78.9474**
17	100	94.7368	**94.7368**	92.1053

Testing o2				
	trainlm	*trainbfg*	*trainscg*	*traincgf*
5	**57.1429**	**71.4286**	42.8571	42.8571
6	0	42.8571	28.5714	**42.8571**
17	14.2857	0	**42.8571**	0

Figure 11. Prediction accuracy of the output Disease

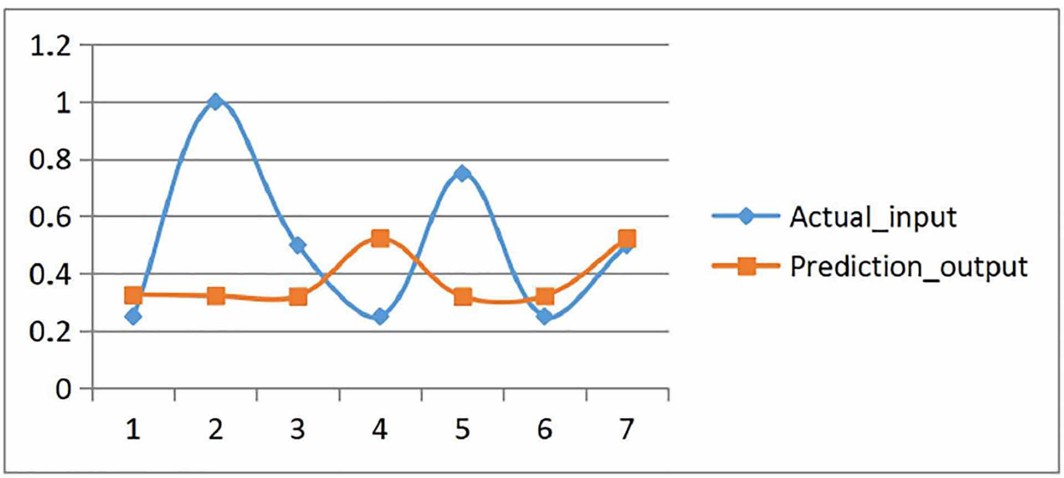
Training algorithm	*tansig*					*logsig*				
	Number of hidden neurons	Epochs reached	Error	Prediction accuracy		Number of hidden neurons	Epochs reached	Error	Prediction accuracy	
				Training	Testing				Training	Testing
trainlm	18	324	0.000987	100%	100%	6	1000	0.0144	100%	100%
trainbfg	5	717	0.0201	100%	71.4286%	18	1000	0.00561	100%	100%
trainscg	15	1000	0.0107	100%	100%	8	1000	0.0196	100%	100%
traincgf	8	1000	0.0192	100%	100%	5	1000	0.0252	100%	71.4286%

Expert System Prototype Evaluation

Researchers have visited zonal agricultural and horticultural research station, Brahmavar from time to time to collect information as represented in Figure 12. Authors discussed with agriculture experts and scientists about the problems related to agriculture diseases. The information is gathered through interactions on the details of diseases and their symptoms of paddy crop. This information was useful and very important for designing expert system to guide farmers. This information helped us to classify different types of diseases in to a set of 24 diseases commonly especially found in coastal areas. Using this information, the farmer can get help for growing their crop and analysing the diseases. The comparison of predicted results and actual results are tabulated in the Figure 13.

Figure 12. Discussion with Associate Director of Research Dr. Hanumanthappa and Pathologist Kavyashree

Figure 13. Comparison of actual results and predicted results of various stages of Plant diseases by using logsig transfer function



FUTURE RESEARCH DIRECTIONS

Authors propose to implement this expert system model using Intel Distributed Python tools. The accuracy in predicting the paddy crop diseases will be computed using various algorithms such as liner regression, decision tree, random forest and K nearest neighbourhood. The best algorithm will be suggested after comparative study of algorithms.

The work on developing mobile apps for farmers is in progress. The farmers can download the app from app store into their smart phones. Farmers can click and upload the images of paddy crop and answer simple questionnaire using their smart phone to get the online advice from expert systems deployed in cloud anytime and anywhere.

CONCLUSION

An expert system is a powerful tool for getting the expert advice at anytime and anywhere by the farmers using tools such as website and smart phones. These expert systems encode and store the knowledge of domain experts. The expert system imitates the thought process of experts in giving advises. Authors have captured the knowledge from experts and the real data from the field study while designing the expert system prototype. This prototype has been experimented with Matlab Toolbox. The results obtained from the expert system prototype are satisfactory. This prototype has been presented to agriculture experts for evaluation.

The summary of the activities carried out are:

1. Collected images of different types of diseases and classified the symptoms pertaining to each disease.
2. The disease impact is classified in to four stages such as preliminary, matured, serious and dead.
3. The Fuzzy rules are designed based on the diseases, symptoms, and stages.
4. The Fuzzy logic toolbox is used managing the rule-base.
5. The surface viewer and the rule viewer tools are used for presenting the results graphically.
6. Some diseases have some common symptoms. So combined rules are designed for such diseases.
7. Data set is divided in Training set 80% and Testing data set 20% of data.
8. The neural networks toolbox is used to train and test the expert system prototype.
9. The prediction accuracy was calculated by using the different transfer functions and training algorithms.

REFERENCES

Agus, F., Ihsan, M., Khairina, D. M., & Candra, K. P. (2019). *Expert System for Rice Plant Disease Diagnosis*. Retrieved from https://www.researchgate.net/publication/331260352

Das, S., & Nayak, L. (2013). Expert System for Agriculture Extension. *Indian Journal of Automation and Artificial Intelligence, 1*(2). Retrieved from http://ijaai.informaticspublishing.com/index.php/ijaai/article/view/32577

Deepthi, M. B., & Sreekantha, D. K. (2017). Application of Expert Systems for Agricultural Crop Disease Diagnoses - A Review. *International Conference on Inventive Communication and Computational Technologies (ICICCT 2017)*. Retrieved from https://ieeexplore.ieee.org/iel7/7966452/7975154/07975192.pdf

Elemo, T. K., & Kumar, M. (2018). Diagnosis System for Disease in Wheat plant using Intelligent Expert System: Design and Implementation. *International Journal of Pure and Applied Mathematics, 118*(24). Retrieved from http://www.acadpubl.eu/hub/

Kannan, P. K., & Hemalatha, K. (2012). Agro Genius: An Emergent Expert System for Querying Agricultural Clarification Using Data Mining. *International Journal of Engineering and Science, 1*(11), 34-39. Retrieved from http://rolliawati.dosen.narotama.ac.id/files/2012/03/fuzzy-toolbox-in-matlab.pdf

Kaur, N., Singh Rddi, N., & Payyail, A. (2013). Review of Expert system based of Fuzzy logic. *International Journal of Academic Research in Computer Engineering, 2*(3). Retrieved from https://www.ijarcce.com/upload/2013/march/4-anand%20nayyar%20-20Review%20of%20Expert-c.pdf

Li, Y., Zhang, W., & Yang, H. (2012). The Application of Expert System and Neural Networks in Crop Growth Management System. *7th International Conference on Computer Science & Education (ICCSE 2012)*. Retrieved from https://ieeexplore.ieee.org/document/6295145/

Peng, Y., Wang, Y., Tang, J., & Li, Z. (2010). Study on rice pest insect and disease diagnosis expert system in Jiangxi province using BP Neural network. *3rd International Conference on Computer Science and information Technology*, 432-434. Retrieved from https://ieeexplore.ieee.org/document/5564128/

Singh, P.K., & Hussain, M.S. (2018). Expert System for Crop Selection in Agriculture: A Critical Review. *International Journal of Advanced Research in Computer Science, 9*(2).

Sreekantha, D. K., Deepthi, M. B., & Puneeth, R. P. (2017). Expert Systems for Disease Diagnoses of Crops - A Study. *XIII Agricultural Science Congress-2017: Climate Smart Agriculture*.

Sriram, N., & Philip, H. (2014). *Expert System for Decision Support in Agriculture, 2014*. Retrieved from http://agritech.tnau.ac.in/pdf/14.pdf

Tatte, M.K., & Nichat, M.K. (2013). Enhancement in Agro Expert System for Rice Crop. *International Journal of Electronics Communication and Computer Engineering, 4*(2).

Chapter 4
Deep Learning and Computer Vision in Smart Agriculture

Shiv Kumar
https://orcid.org/0000-0003-4792-9207
Netaji Subhas University of Technology, India

Agrima Yadav
Netaji Subhas University of Technology, India

Deepak Kumar Sharma
https://orcid.org/0000-0001-6117-3464
Netaji Subhas University of Technology, India

ABSTRACT

The exponential growth in the world population has led to an ever-increasing demand for food supplies. This has led to the realization that conventional and traditional methods alone might not be able to keep up with this demand. Smart agriculture is being regarded as one of the few realistic ways that, together with the traditional methods, can be used to close the gap between the demand and supply. Smart agriculture integrates the use of different technologies to better monitor, operate, and analyze different activities involved in different phases of the agricultural life cycle. Smart agriculture happens to be one of the many disciplines where deep learning and computer vision are being realized to be of major impact. This chapter gives a detailed explanation of different deep learning methods and tries to provide a basic understanding as to how these techniques are impacting different applications in smart agriculture.

INTRODUCTION

With the rapid increase in global population, challenges related to food supplies are becoming a huge concern. UN reports [26], project upwards of 9.7 billion people on the planet by the year 2050 and according to The UN Food and Agriculture Organization, feeding this expected population would require augmenting the current food production by approximately 70% (Jin, Fu & Zhang, 2014). Also, it is

DOI: 10.4018/978-1-5225-9632-5.ch004

important to keep in mind that millions of people around the world are dependent on agriculture for their livelihood and suffer from different issues impacting the yield of their crop which are not limited to soil incompatible plantation of crops, fertilizer overuse, weeds, improper management of yield and plant diseases. Traditional methods are also known to cause degradation to the environment which in turn acts as a contributing factor to the reduction of productivity in the future. Employment of labor in tedious tasks ought to be automated in order to better manage costs incurred during the maintenance of the crop. This rise in demand coupled with the inability of conventional methods to meet the current demands and necessities of food supply, are a clear indication that better and more efficient methods need to be devised in order to match the expected needs of the future, else a global food crisis will be inevitable. These methods are being understood to be a combination of productive traditional practices and the innovations of agronomists and agricultural engineers, leading to the growth of a field known as Smart Agriculture. Smart agriculture enables better monitoring, analysis and understanding of the complex agricultural ecosystem, bringing together different technology-based approaches to the diverse field of agriculture. It would be an understatement to say that smart agriculture is the key to a secure and sustainable future, and it is the sub-fields of Agricultural Disease Management and Crop Prediction which are expected to make the greatest impact.

The first and most important aspect that technology in relation to smart agriculture covers, is the dissemination of the right information to the farmers at the right time. When crops are grown in remote, distant farmlands, the services of an expert may be hard to get. In such cases, using a remote application which can provide the necessary information can prove extremely beneficial. Providing accurate weather condition data and weather forecasts shall minimize crop loss resulting from harsh weather conditions. Using images of leaves, an accurate diagnosis of probable plant disease can save the entire crop from an infestation and even recommend the right type and amount of pesticide. Using an entire spatial image of a field taken from a drone or an Unmanned Aerial Vehicle (UAV), the growth of weeds can be detected and controlled amounts of herbicides can be used, minimizing the environmental harm caused by uncontrolled pesticide use, allowing for efficient management of agricultural diseases. The spatial images taken by the drone can also be used to predict healthy yield from the crop, as well. Throughout the supply chain, food is wasted due to problems in harvesting, storage, packaging and transportation. A fruit counting and yield production algorithm for commercial purposes can help growers plan the manual labor requirement for all the tasks and make necessary packing and storage arrangements before their sale, thereby highlighting the importance of accurate yield prediction.

Two different but linked domains contributing heavily to groundbreaking innovations in the field of smart agriculture-based work in management of agricultural diseases and yield prediction are that of Deep Learning and Computer Vision. Deep Learning is a part of a family of different machine learning algorithms which involves the training of multiple layers of perceptrons, modelled based on the functioning of the network of neurons in the human brain; sometimes also referred to as a neural network. These artificial networks can be fed text, images, videos or audio as input and they learn complex relationships between the said input data. The learned relationships are later used to perform different kinds of tasks in the required series of domain specific operations and applications. Deep Learning finds use in the fields of image and pattern recognition tasks, face recognition on social media applications, speech recognition tasks like the voice control built into smartphones, speakers and other electronic devices and even sentiment analysis tasks that involve understanding the class of opinion behind written text. Computer vision on the other hand is a specific field which aims to provide machines with a similar if not better sense of vision as compared to humans, i.e. being able to infer and extrapolate different forms of information from a single or sequence of images. Agricultural applications encompass a lot of operations related to images, such as Image classification, object classification and localization, land cover classification, obstacle detection, plant pheno-typing, fruit counting, etc. Hence being able to extract and infer useful information from images provides encouraging results in all sub fields. Computer Vision has been hugely driven by deep learning techniques since the rise of Alex-net in the ImageNet Large Scale Visual Recognition Challenge in 2012 (Krizhevsky, Sutskever & Hinton, 2012). Deep Learning is what provides the inference capability in Computer Vision, while Image Processing coupled with data collection techniques form the other half of the spectrum. To understand how, the combination of deep learning and computer vision is impacting the field of smart agriculture driven management of agricultural diseases and yield prediction is the main objective of this chapter.

This chapter tries to explain in as much detail as possible, the basic structure of the machine learning process followed by the need for deep learning approaches in the presence of conventional machine learning algorithms which have lesser computational and resource requirements; the concepts that form the basic building blocks of Deep Learning - Artificial Neural Networks, Convolutional Neural Networks followed by the concepts of Recurrent Neural Networks. Next is a discussion about the major applications of these networks in Smart Agriculture, which directly or indirectly affects the sub-fields of Agricultural Disease Management and Yield Prediction, wherein the state of the art in Deep Learning for the specific application has been highlighted. Different frameworks and hardware employed in these applications have been mentioned, to give the reader a better sense of which technologies are being used where and when. The chapter ends with a brief conclusion. The chapter also provides recommendations for subsequent steps to be taken after a thorough understanding of this chapter.

This chapter should be regarded as an introduction to the vast field of Deep Learning and Computer Vision in Smart Agriculture and aims to be a good first resource for a beginner in the field. The authors advise the readers to consult the resources mentioned along with different topics in the future sections for a better understanding of the field.

BACKGROUND

Basic Structure of Machine Learning and the Need for Deep Learning

Any machine learning problem requires the approximation of a model which is *a mathematical equation using which the system generates predictions or decisions for the given problem, which could be either classification or regression, for a certain input, represented by either a feature or a set of features*, based on the given dataset (Bishop, 2006; Kosa et al., 1996). These models are approximated by assigning a coefficient to each feature in the feature set generalized over the dataset and the models can be linear, non-linear, tangents or even sigmoidal equations, depending on the complexity of the feature set. Here, datasets are a group of samples represented by a set of distinguishable features (together referred to as a feature set), collected by individuals working in the respective field, having undergone a certain degree of cleaning and pre-processing steps, before being ready to operate on.

Machine learning has majorly two branches, i.e. supervised and unsupervised learning. The process of developing a model in supervised learning for the given dataset majorly involves two phases namely training and testing. Training is the phase where the system learns the model for a subset of the dataset which is then tested on the remaining data to evaluate its performance on data that has never been encountered by it before. This is done by comparing the predictions or decisions made by the model with the actual true values (referred to as ground truth in some texts of the literature concerning machine learning) for the corresponding feature sets associated with the problem which are also present in the dataset.

When a model performs well on the training subset but not so much on the testing subset, then it is said to be learning for each input value (analogous to memorizing and not understanding), rather than generalizing. This is unfavorable due to the noise attached to each input, which is also learnt during the process, and should be avoided as much as possible. Therefore, the model should generalize over a dataset. This phenomenon is called overfitting and is something any model should avoid at all costs. There is another concept which is in complete contrast to overfitting i.e., the model not being able to reach a certain level of accuracy for training as well, due to poor approximation of the model called underfitting, which can easily be avoided by coming up with better models.

To check the performance of a model, the error or loss function is used, which is an indicator of the deviation of the model from the true values of the dataset. The minimization of the error function is one way to move towards a better model. This process of moving towards a better model by gradual reduction in the error function value is called gradient descent (Goodfellow, Bengio & Courville, 2016) and works by updating the coefficients given to different features. Initially random weights are taken, and then the weights are updated by subtracting from them, the derivative of the error function with respect to the specific weight, times the learning rate, as shown in Equation 1.

Learning rate is introduced to be able control the rate of update for weights. Very large or very small updates add instability to the learning phase which is known to be undesirable in any machine learning algorithm and hence learning rate is used to control the extent of change in weights.

Equation 1: Updation rule for the i^{th} coefficient or weight of the model

$$w_i = w_i - \alpha \frac{dE}{dw_i}$$

Unsupervised learning on the other hand, is majorly concerned with identifying patterns using the model approximations in the dataset and then separating the samples into logical groups based on the approximated model. This is done by identifying the similarity between the different features of different samples in the datasets, and then separating them on the basis of some similarity metric. This understanding is enough for the readers to understand the content related to unsupervised learning in this chapter. The readers can refer(Gerven & Bohte, 2017) for a more detailed understanding of the topic.

Most real world-based datasets are not linear, and need more complex equations to be able to operate efficiently. Conventional machine learning algorithms, generally require addition of dummy features i.e. quadratic or similar dependencies to other features of the feature set, hand engineered by the researchers to come up with these non-linear or more complex equations since the algorithm is based on a linear equation to begin with. This process is heavily dependent on the researcher's knowledge of the dataset and may not always be correct. It may also happen that the best features for the model to train on are simply not taken into account due to some bias or misconceptions on the expert's side. In simple words, operating on dummy features to find complex models is a tough task and researchers look for techniques and algorithms which are able to find appropriate decision boundaries without involving dummy features to the feature set.

This is where the concept of neural network(s) is of utility. They are able to develop complex models for the problem without having to add dummy features into the dataset, by forming composite unions of different individual models, and have given state of the art results when compared to their contemporaries. For a more detailed understanding of this process, the authors recommend the readers to refer (Goodfellow, Begnio & Courville, 2016).

The following section deals with understanding different types of neural networks i.e. ANNs (Artificial Neural Networks), CNNs (Convolutional Neural Networks), RNNs (Recurrent Neural Networks) which are the major basis of understanding deep learning and how it impacts different areas of smart and precision agriculture.

Artificial Neural Network

A neural network (Gerven & Bohte, 2017), as shown in Figure 1. can be regarded as a group of individual systems or models which when combined in a layered architecture or structure, can be used to define a composite system with a very complex boundary for a given dataset, while also avoiding the model from overfitting.

Based on the structure of the biological brain, the individual systems are called neurons (or units) which are connected to other systems in different layers. The initial layer of this architecture is called the input layer, and the final layer involved in giving the output for the problem is called the output layer. The layers in between the input and output layers are referred to as hidden layers. The input layer of the network has the features of the dataset as its units. The units in the hidden layers, and the output layer form separate individual models depending on the weight assigned to the respective units connected to it from the previous layer in the model.

For example, any unit in a hidden layer takes in as input the output of all units in the previous layer of the network. In the unit of the hidden layer, a weighted sum of all the different inputs is taken and the weighed sum is then passed through a non-linear function called the activation function. This activation function is the source of non-linearity in the model.

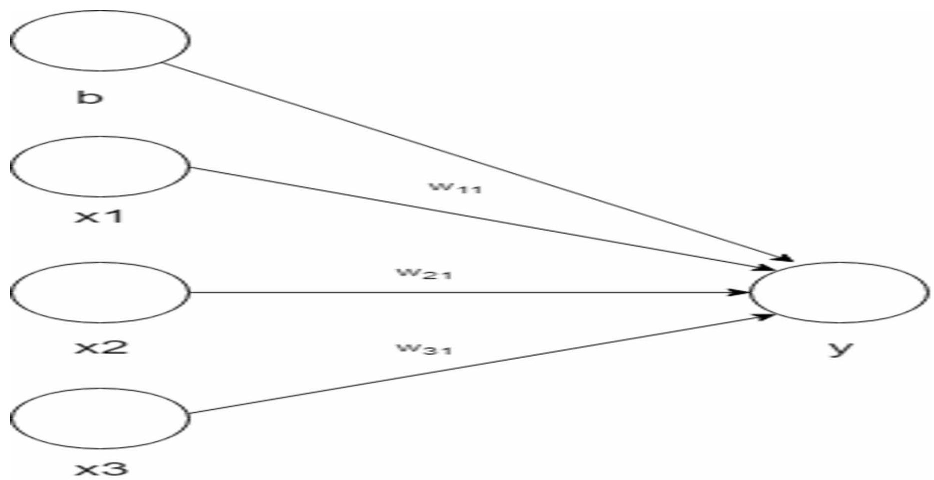
As the network gets deeper and deeper, the units in the deeper hidden layer take in as input more complex models from the previous layers, and further output, more complex models, until the output layer is reached. In the output layer, the net result of all models, is taken to develop the cumulative model for the system, which decides the output of the network.

Popular activation functions used in deep learning are SoftMax, ReLU, ELU and Tanh. More information for the same can be found here (Gerven & Bohte, 2017).

The process of outputting for an input feature set is called forward propagation and requires a single traversal of the network. These outputs require weights on the connections between the neurons as mentioned before. Outputs are evaluated for the inputs using random weights at the beginning. Then the amount of error being induced in the network due to the random weights using a loss or cost function, is evaluated. If the error being induced in the network can be reduced to a satisfactory level, then it can be said that the system is optimizable.

The error is reduced using the process of back propagation. This process involves the updating of weights of the network by evaluating the dependency of the error on the weight, for which the derivative of the error function with respect to the weight is calculated (Nielson, 2015). Once this dependency is evaluated, the weight is updated by subtracting the derivative times the learning rate as highlighted in Equation 1.

Figure 1. Artificial neural network



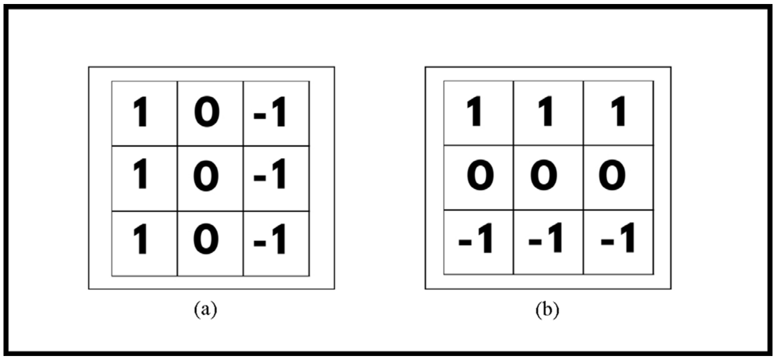
Here, the x1, x2, x3 represent the features of the dataset, whereas the w_{ij} represent the weight of the connection between i^{th} unit in the current layer and j^{th} unit in the next layer. The unit marked as 1, along with b weight represents a bias element to add a constant term to the model.

Convolutional Neural Networks

While working on images, the type of processing discussed in artificial neural network will treat each pixel as a different feature, and will not account for the correlation between different pixels i.e. it does not account for how different pixels combine to form different shapes and curves. The human brain on which the ANN is based, accounts for these correlations and forms a suitable boundary to recognize objects with a combination of pixels processed together as is done in the visual cortex of the brain. A computer accounts for these correlations using convolutional neural networks. It is based on the principle that correlation can be attributed to a combination of pixels i.e. being able to form a super pixel from a combination of sub pixels. To form this super pixel, it is necessary to find a logical combination of the sub pixels. One such combination can be the weighted sum of these values, since the contribution of one pixel will be more than the other in deciding the value of the super pixel. The convolutional neural network should be able to learn these weights itself as it does in ANN. These weights should be assigned in such a way that it is able to identify patterns and shapes in the image. This is the basic concept and strength of Convolutional Neural Networks i.e. given an image it should be able to output the correlated features in an image which can then be fed into an ANN which can then classify the images as per requirements of the application at hand.

Convolutional Neural Networks are essentially deep artificial neural networks that find applications in the fields of Object Detection, Object Detection and Localization, Image Classification, Scene Labelling, Video Activity Recognition, etc. Convolutional Neural Nets are well qualified for image processing tasks.

Figure 2. Filters for (a)vertical and (b)horizontal edge detection



The Convolution Operation

A grayscale image is represented as a two-dimensional matrix. If a filter is applied to the image, the input image is said to be convolved with the filter matrix. The filter may be a vertical edge detector - to detect all vertical edges in the given image - for example, a 3 x 3 vertical edge detection matrix is shown in Figure 2 (a) or a horizontal edge detector – to detect all horizontal edges in the given image – for example, a 3 x 3 horizontal edge detection matrix is shown in Figure 2 (b).

It can also be a filter to detect edges at 45 degrees or 70 degrees.

On an input matrix of size 4 x 4, the convolution of the 3 x 3 vertical edge detection matrix is illustrated below in Figure 3.

Figure 3. Convolution operation over the given matrix

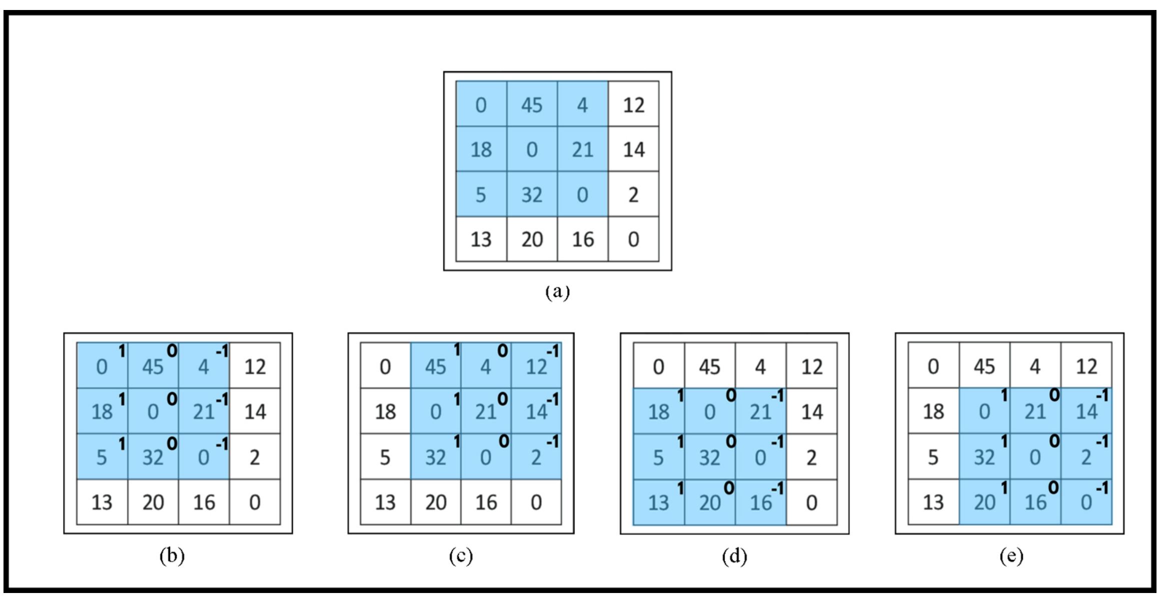
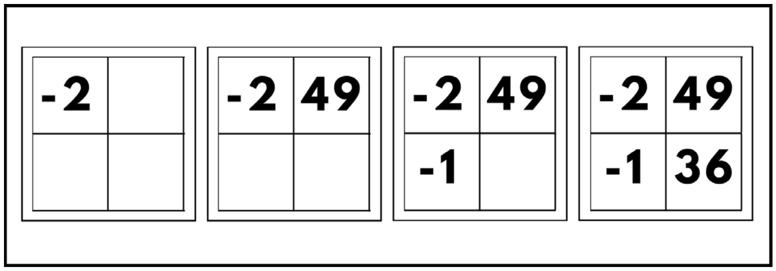


Figure 4. Corresponding results of the convolution operation



The steps involve multiplying each element of the filter matrix (shown in superscripts) with the input matrix elements that it overlaps with and storing the sum as the result in the corresponding position.

In step 1, the result is obtained as 0(1)+45(0)+4(-1)+18(1)+0(0)+21(-1)+5(1)+32(0)+0(-1) which is equal to -2, which forms the first element of the result matrix. Similarly, the result has been computed for the other three steps as shown in Figure 4.

For a greyscale input image of size n x n convolved with a filter matrix (or convolution matrix) of size f x f, the resulting image is a matrix of size: (n-f+1)*(n-f+1). Applying the formula on the above example where n=4, f=3, the resulting matrix is of size (4-3+1) x (4-3+1) or 2 x 2.

Padding

As observed in the formula given above, the output image keeps shrinking as filters are applied, disregarding information from the edges of the input image. In such cases, it may be useful to pad the image with an additional border of say *p pixels* (as shown in Figure 5) so that the output image is now of the dimensions: (n+2p-f+1*n+2p-f+1). This allows taking all pixels into consideration equal number of times while evaluating the convoluted image.

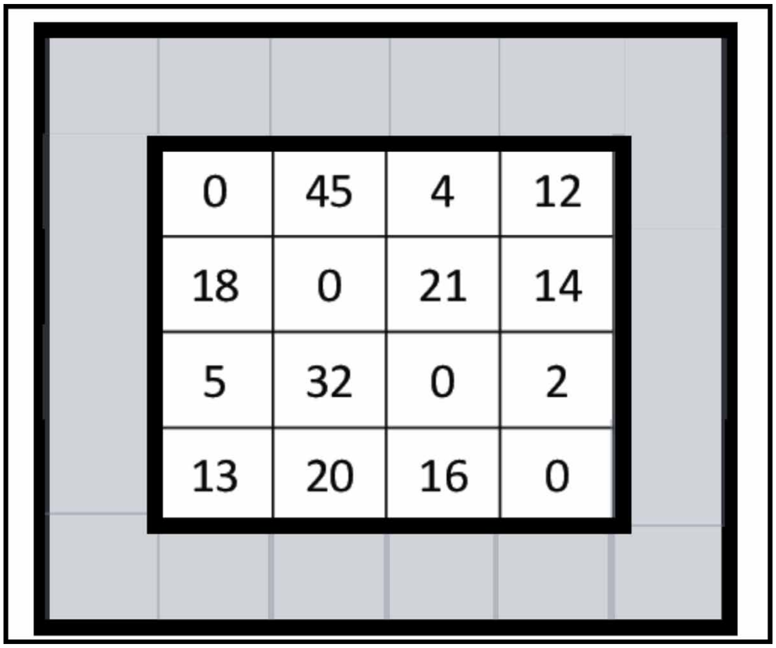
Based on the padding, convolutions may be classified into:

1. **Valid Convolutions:** No padding is done on the input image (p= 0). The n x n image is convolved with an f x f filter to give an output image of size (n-f+1)*(n-f+1).
2. **Same Convolutions:** Padding is performed in a manner that the output size is the same as the input size. So, n+2p-f+1, equals n, and therefore p here should equal to (f-1)/2 and since filters are usually odd number matrices for most practical applications, the value of padding comes out to be a whole number.

Stride

In the above example, for every subsequent computation, the filter shifts by one column. For such cases, the stride (denoted by s) is 1. Given a stride of 2 (s=2), the filter shifts by two columns in the next computation. So, for an n x n image, with a padding of p pixels, an f x f filter and a stride s, the dimensions of the final image will be *floor*{((n+2p-f)/(s+1)) * ((n+2p-f)/(s+1))}. For computation, the filter must lie within the image or the (image + padding)

Figure 5. Padding applied to a matrix



Convolution Over Volume

For most practical applications these days, the input to a convolutional neural network is a multi-channeled image comprising of three channels - RGB. For example, the dimensions of the image may be 4 x 4 x 3, where 3 corresponds to the Red, Green and Blue color channels. While the height of the image is 4 pixels and the width are 4 pixels. This input image can be visualized as a stack of three 2D images of size 4 x 4. It is also important to note that the number of channels in the filter are taken to be the same as the number of channels in the input image.

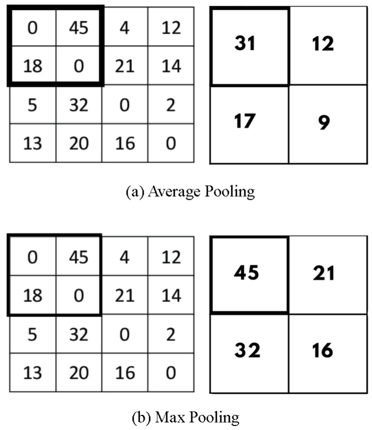
If dimensions of the image are n*n*nc and dimensions of the filter are $f*f*f_c$, then the resulting image is of the dimensions: $(n-f+1)*(n-f+1)*n_c$. (*or number of channels (n) of the next layer, which is just equal to the number of filters applied*).

One Layer of a CNN

The image at hand is convolved with n filter matrices and the output is summed with a bias (for each entry in the matrix). This is passed into an activation function like ReLU, resulting in the final matrix. All these operations together form the convolution layer of CNN.

The Convolution layer, pooling layer and fully connected layer constitute the basic building blocks of a CNN. Other than convolution layers, Convolutional Neural Networks often use pooling layers to reduce the spatial size of representation in order to reduce the number of parameters and computation in the network. The popular types of pooling are:

Figure 6. (a) Average pooling and (b) Max pooling



1. **Max Pooling** (as shown in Figure 6(a))**:** For each of the grids highlighted in the vector below, the maximum value is selected as output. Pooling does not have to learn parameters. Instead, it has a fixed set of hyper-parameters and fixed calculations. On the other hand, a large number of parameters exist in fully connected layers.
2. **Average or Mean Pooling** (as shown in Figure 6(b))**:** The input is divided into smaller regions and average values of each region are computed.

Dropout Layer

Convolution networks also employ a layer by the name of dropout layer to increase the randomness factor in the network to avoid overfitting. This layer exists between the dense and output layer, allowing only certain units to go through depending on some probability factor.

Recurrent Neural Networks

For many problems, input data is sequential or has a factor of context associated with it. For example, sequence of words, audio signals or videos. But till now, the discussion has consisted of models that do not take this factor into account i.e. the models are provided input in one go and they generate outputs in one go without taking into consideration the temporal context between the features. Examples of such problems include music generation, named entity recognition, machine translation, speech recognition, video activity recognition, etc.

Depending on the application, both the input and the output can be variable sized, with a sequential nature to them. The basic idea behind RNNs is that when a unit receives some input feature, it stores the information corresponding to it and that information is used while processing the next input feature and all future input features i.e. RNNs have a sort of memory element associated to it which allows the network to keep track of a context while processing features encountered at different time steps. RNNs are famously known to be able to learn shared features across different positions in the sequence. There are three different types of units that are used in RNNs with different capabilities and varying complexities i.e. the basic unit, the Gated Recurrent Unit (GRU) and Long Short Term Memory (LSTM). GRUs and LSTMs are based off the basic unit of RNNs, and these basic units are instrumental in understanding RNNs. RNNs generally operate on vectors.

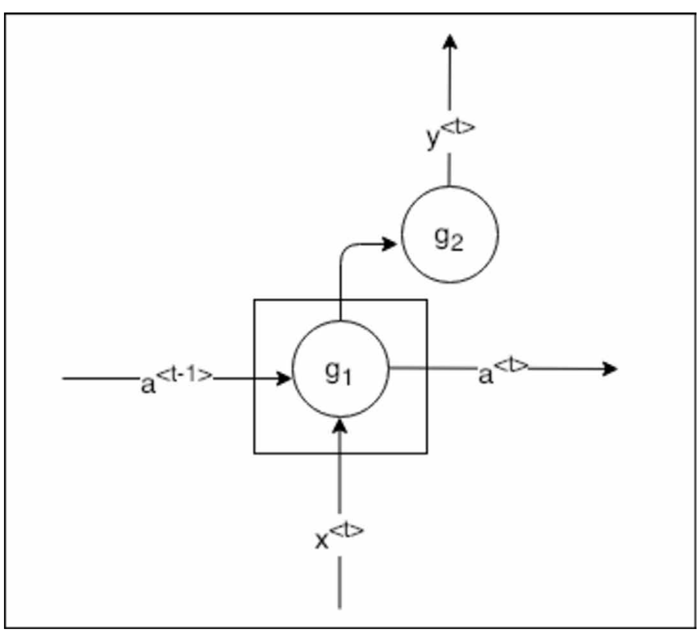
The Basic Unit

The features are inputted to the basic unit (shown in Figure 7) in a sequential order, where an activation(represented by g_1) is calculated for each feature at different time steps in the unit, by applying an activation function on the multiplication of the input feature (represented by $x^{<t>}$) and the weights summed to the activation from previous time step (represented by $a^{<t-1>}$). The output for each time step in the unit may be calculated by applying another activation function(represented by g_2) on the activation obtained by using the unit. In some applications the output might be taken after processing all features of the input, and some might output parallel to the input features being processed. This output is referred to as $y^{<t>}$ henceforth. To maintain consistency an activation is also added at the 0^{th} time step which has a vector of zeros. The calculation of the activations as well as the output at each time step are shown using Equation 2 and Equation 3.

Equation 2: Activation for the next time step

$$a^{\langle t\rangle} = g_1\left(w_{aa}a^{\langle t-1\rangle} + w_{ax}x^{\langle t\rangle} + b_a\right)$$

Figure 7. Basic RNN unit



Equation 3: Output corresponding to each time step

$$y_p = g_2\left(w_{ya}a^{\langle t\rangle} + b_y\right)$$

An obvious weakness of such an approach is that it only takes into consideration features of previous time step into consideration while making a decision for the current input feature which might not always be the best approach for a problem. For example, named entity recognition, as explained below.

He said, “Rose is the best player in the NBA”.
He said, “Rose is the most economical flower in agriculture”.

The process of back propagation here also deals with tuning the weights so as to minimize the loss function as much as possible. The loss function is taken as the aggregate of all-time steps. Another problem that RNNs face is of vanishing gradients. Here the gradient refers to the gradient of the loss function with reference to the weights, calculated at the time of back propagation, which are used to update the weights. Vanishing gradients refers to the gradient of the loss function with reference to weights in the earlier layers becoming very small, due to the involvement of derivatives which can take very small values, and as discussed before, the updating of individual weights involves product of different derivatives. These can get very small for the earlier layers, and the weights don’t really change much and their hit on the error doesn’t reduce, making it feel as though the network is stuck. This is avoided by the use of other type of units of RNN i.e. GRUs and LSTMs.

Gated Recurrent Unit

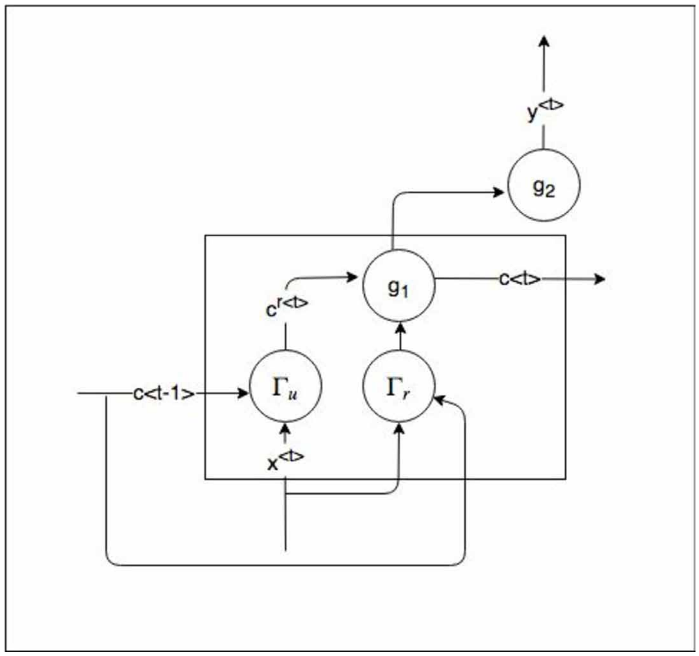
GRU (Namin et al., 2017) (shown in Figure 8) is a modification to the basic unit of RNN, which makes it much better at capturing long range connections in the sequence. An example of long-range connections is as shown below.

The seed, which already grew — is harvest ready.
The seeds, which already grew — are harvest ready.

There could be many words between slept and the helping verb before awake, so it could be important to remember the context that the baby was singular in the first sentence whereas it was plural in the second example to be able to make prediction of possible words in different language models.

GRUs have a memory unit referred to as c represented using $c^{<t-1>}$ and $c^{<t>}$ depending on the time step involved which, as the name suggests, provides memory to remember context to the RNN units. In GRUs the content of the memory cell and the activation are same. Like in the example considered before, 1 can be stored in c if the word to take into context is plural or 0 if it is singular, and till it does not go out of context, it can be maintained in the memory cell, and once it does (depending on the rules defined by the designer of the model), it can be considered to overwrite it with a different context depending on the local active context. That is the basic principle behind GRUs. At each time step the model considers overwriting, the value of the memory cell with a value of $c^{r<t>}$. $c^{r<t>}$ refers to the candidate value that may replace the current value of the memory cell.

Figure 8. GRU unit



The GRU consists of two gates (represented by Γ_r and Γ_u) which decide whether the candidate value will actually replace the current value of $c^{<t>}$, or not. The equations involved in GRUs are as follows:

Equation 5: Calculation of the update factor using the Update Gate

$$c^{r\langle t\rangle} = \tanh\left(w_c\left[\Gamma_r \times c^{\langle t-1\rangle}, x^t\right] + b_c\right)$$

Equation 6: Calculation of the remember factor using the Remember Gate

$$\Gamma_u = \sigma\left(w_u\left[c^{\langle t-1\rangle}, x^t\right] + b_u\right)$$

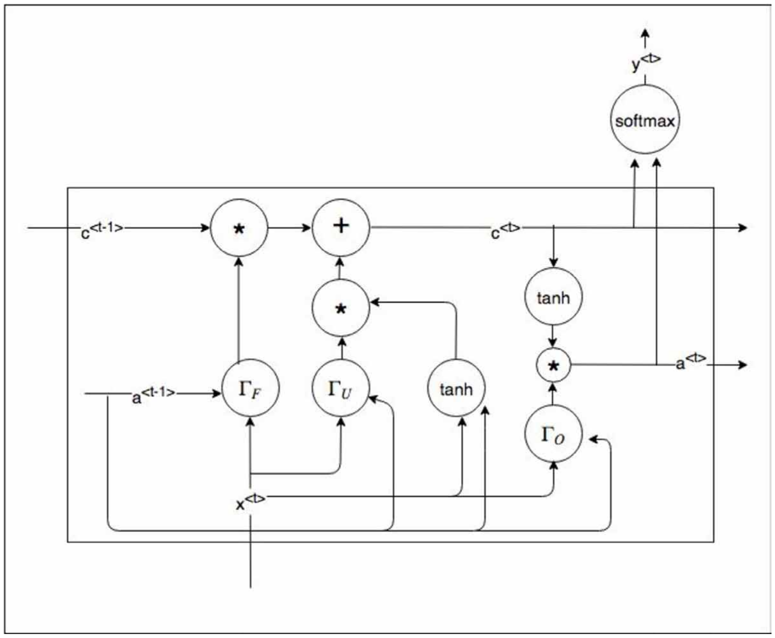
Equation 7: The calculation of the value of the memory unit

$$\Gamma_r = \sigma\left(w_r\left[c^{\langle t-1\rangle}, x^t\right] + b_r\right).$$

Long Short-Term Memory Unit

The LSTM[5][19](as shown in Figure 9) can be considered an extension of the GRU where the activation(represented by $a^{<t-1>}$) and memory cell(represented by $c^{<t-1>}$) values are not taken to be same. The unit consists of two other gates other than the update gate (represented by Γ_u) i.e. forget gate and the activation output gate (represented by Γ_f and Γ_o). The update gate provides weight to the new candidate

Figure 9. LSTM unit



value to be able to replace the older value, whereas the forget gate gives weight to the old candidate whether it should be forgot or not. Their dependency together decides the value of the memory cell in the current time step. The output gate operates on the activation from the previous time step, which is used in deciding the activation for the next time step. Here the candidate value is calculated using the activation value as shown in the following equations.

Equation 8: Calculation of the candidate value for the memory unit

$$c^{r\langle t\rangle} = \tanh\left(w_c\left[a^{\langle t-1\rangle}, x^t\right] + b_c\right)$$

Equation 9: Calculation of the update factor using the Update Gate

$$“_u = \sigma\left(w_u\left[a^{\langle t-1\rangle}, x^t\right] + b_u\right)$$

Equation 10: Calculation of the forget factor using the Forget Gate

$$“_f = \sigma\left(w_f\left[a^{\langle t-1\rangle}, x^t\right] + b_f\right)$$

Equation 11: Calculation of the output factor using the Output Gate used in calculation of activation

$$“_o = \sigma\left(w_o\left[c^{\langle t-1\rangle}, x^t\right] + b_o\right)$$

Equation 12: The calculation of the value of the memory unit

$$c^{\langle t \rangle} = \text{“}_u \times c^{r\langle t \rangle} + \text{“}_f \times c^{\langle t-1 \rangle}$$

Equation 13: The calculation of the activation for the current time step

$$a^{\langle t \rangle} = \text{“}_o \times c^{\langle t \rangle}$$

Bidirectional RNNs

Bidirectional RNNs solve the issue where the model was operating on current input by taking into consideration only the previous time steps' input features. The activation from both the previous and future time steps are taken into consideration while deciding the activation for the current time step unit i.e. the model propagates from both directions. The major constraint with bi-directional RNNs is that the entire sequence is required before operating, which might not be appropriate for real time applications of RNN.

DEEP LEARNING IN SMART AGRICULTURE

This section provides detailed explanation of a few applications which impact the fields of Agricultural Disease Management and Yield Prediction directly or indirectly. These applications were chosen keeping in mind the degree of their impact in respect to the scope and limits of the book as well as the impact of deep learning to these applications.

Plant Disease Detection

For identifying plant diseases, the services of professional agriculture engineers can be employed but if the crop grows in a remote area, professional services may not be readily available. For such cases and for early detection of plant diseases before they spread to the entire crop, remote monitoring through either smartphone applications, Unmanned Aerial Vehicles (UAVs) or other autonomous systems may be performed. Detection and cure of the diseases in the early stages significantly reduces the agricultural cost of production, while also increases the productivity of the crop, and hence of significant importance to Agricultural Disease Management.

Pathogen infected leaves may have spots, patches, rings, vein banding or the appearance of a powdery mildew. In the deep learning models developed for the detection and diagnosis of plant diseases, convolutional neural networks were employed on images of leaves of several healthy plants along with diseased plants. As discussed before, CNNs are suitable for recognition of patterns in a large dataset of images. Several CNN systems have been developed - an example of which is a neural network to successfully identify legume species based on the patterns in the leaves' veins. Several architectures to identify plant diseases use an open dataset of fourteen different plants. Although the techniques had a high rate of success, but the major drawback identified was that the dataset only included photographs of leaves in the laboratory setup and not in the real field conditions with soil, fruits, multiple leaves, and other noise elements in the background.

In Kosa et al. (1996), the authors use a database of photographs captured in both the laboratory environment as well as real cultivation field environment. Five different architectures studied in the work include AlexNetOWTBn, VGG, Alex Net, GoogLeNet and Overfeat.

For the training and testing of the models, an open dataset of 87,848 images of leaves of both healthy and infected plants was used for a total of 25 plant species. It had 58 different classes with each class denoting a {plant, disease} pair and also healthy plants with no disease. 37.3% of these were images from the field. A random split of 80/20 on the images led to the training set and the testing set. The final performance of the models showed a success rate of 99.49% on the testing dataset in case of the AlexNetOWTBn with a time 6647 seconds per epoch and the highest success rate of 99.53% for the VGG with a time of 7034 seconds per epoch. Total training time for the model was about 5.5 days. The results also show that models achieve better performance when trained on field images and tested on lab images but if trained entirely on experimental lab images and tested on field conditions, success rate is significantly reduced.

With the computational power of 2ms to classify an image, it can be possible to integrate the disease detection application into smartphones. Once the farmer knows the plant disease, with expert advice, he can be made aware of the right pesticides to use and their quantity, eliminating the problem of overuse of these chemicals in the cultivation.

A Windows phone application capable of detecting vineyard plant diseases with an accuracy of 90% has been described in Kosa et al. (1996). The system isolates from images of fruits and leaves, the spots and the analysis are based on their size, number and color. Using the GPS location of the device and relevant weather data along with humidity and temperature, the occurrence probability of the particular disease can be assessed. It was implemented in Visual Studio 2015 using the Silverlight library and displays the three most probable diseases with the given conditions.

Weed Identification

The presence of weed in a crop field, leads to an increased irrigation requirement and cost of agriculture. They compete with the crop plants for sunlight and nutrition; affect the purity of crops by cross-pollination and act as a breeding ground for harmful pathogens. That not only affects the yield but also, leads farmers, to make inaccurate predictions in respect to the yield, which is clearly very harmful. Cultural and modern practices in weed management include altering the row space between crops, growing the varieties that compete well with weed and spraying of pesticides and herbicides over the entire field, leading to herbicide pollution and increased agricultural cost. The concept of precision agriculture encompasses allocation of the correct amounts of herbicides, restricted to only the places where spraying them is necessary and at the right times. The challenging problem for Deep Learning methods in weed identification is the fact that in appearance, weed looks very similar to the crop grown when photographed through Unmanned Aerial Vehicles or UAVs. UAVs or drones can obtain photographs of the complete agricultural field at a low cost with high resolution.

The method proposed for identification of weed in Bah, Hafiane, and Canals (2018) is executed in three stages. First, crop lines are detected and interline weeds are identified. In the subsequent phase, interline weed forms the dataset for training. Finally, Convolutional Neural Networks are employed on the dataset of UAV images to detect weed and crop in a fully automatic manner of learning.

The assumption to locate weeds is that plants, depending on their variety, are grown in rows with a fixed space between them. Generally, outgrowth that does not align with the rows are considered as weeds, more specifically known as inter-line weeds. Using this property, an unsupervised model is designed, that is, without the requirement of training data.

In order to detect the main line of each row of crops, the Hough transform was used to highlight the alignment of the pixels, one of the most commonly used methods for detection of lines and pre-processing is done to remove soil or shadows from the images. Residual Network of the ResNet architecture was used in the cited work, a CNN architecture developed in 2015. It had 18 layers according to the size of data. ResNet proved to be superior in terms of results over other architectures such as the VGG13 and the AlexNet in the ImageNet Challenge. The starting point of the research was with the features learned on the ImageNet dataset. From this pre-trained network, the soft max layer was truncated and replaced with a new soft max layer relevant to weed identification, using two categories - crop and weed. Trials were conducted on images of spinach farms and bean farms, acquired by a camera of 36 megapixels at a height of 20 meters. The field was divided into two different parts - one for collection of training data and the other for collecting test data.

Plant Phenotyping

Plant phenotyping is another area where deep learning techniques such as CNNs have seen major application in smart agriculture and precision breeding which encompasses the ability to identify the genus of a plant using its different visual features or traits (phenotypes) such as height, leaf area, roundness, number of leaves, color of leaves etc. This technique is used to select superior variants of a genotype which might help in improving the efficiency in production methods. This application allows, farmers to identify if the genus involved in the yield is growing in the expected manner, which affects the prediction of yield, while also acts as an indicator whether the genus might be mutilated by a disease which might impact the entire yield. Initial work in this area was based on hand crafted features specified by expert biologists, later passed to classifiers to find the most probable genotypes. Apart from not being robust to different datasets and different conditions or experiments, this technique may also suffer from incorrectly measured or missing features, that might be of prime importance. These problems with hand crafted features have been worked upon using CNNs, which extract illustrative features from the sample images at the time of training of the model. Since feature extraction and training is done together, the model finds features that minimize the loss function. But it has been observed that CNNs are not appropriate for dynamic or temporal problems as visual features of different plants may give similar values over different time periods. Therefore, different machine learning practices and experiments dependent on static image samples such as, CNNs which classify similar seeming plants (possibly due to consideration of independence of different time periods) into the same genus, can be avoided by taking time periods also into consideration. Not taking the temporal factor into consideration also leads to incorrect classification of same genus plants since visual features of the same plants might be different at different stages of growth for a plant as well. This is a perfect use case for Recurrent Neural Networks (RNNs), which are able to encode temporal dependencies also into the model. In Namin et. al. (2017), the authors propose a framework composed of a CNN (Alex net in their case) (explained here) to find visual features accompanied by an RNN with LSTM cells (explained here) to recognize time dependent factors in genotype classification.

Their dataset is built using sequences of images of plants on different days of the growth process, broken down to separate images to build a CNN compatible dataset, which is further extending using data augmentation techniques like rotating each image by 90, 180 and 270 degrees around the center. Their process of training the CNN is based on stochastic gradient descent method in batch size, of 32 with learning rate of 0.001, and also involves fine tuning (explained here) on Alex net pre trained on ImageNet. They chose Alex net over other networks such as Resnet or VggNet, since it has fewer weights to be tuned, which suits the constraint of a limited dataset as faced in this situation. After the CNN is trained, the output (deep features of the plant) of the last fully connected layer which was fed into a classification layer in previous approaches is instead fed into a LSTM. LSTM units of the RNN, also trained in batch size of 32, with 256 hidden neurons, are used to help the system model the sequences of deep features of the plant based on the information of the entire sequence. This approach had another layer of LSTM which enhances the ability of the model to learn complex patterns in the sequence, improving the accuracy in the process. This model gives a classification accuracy of 93% as compared to 76.8% of the method where only CNN is employed, proving the temporal dependency of the problem. These results are a major uplift from the accuracy reached on using hand crafted features using SVM which report a classification accuracy of 60.8% only.

These methods will allow as mentioned before to increase efficiency of plant production as mapping can be done for superior genotypes, for certain conditions. This can be of use for plantation in matching climatic conditions to maximize yield.

Fruit Counting

To make an estimation of yield and for orchard management, fruit counting is an important application for farmers or other commercial growers who can then make accurate estimations for their business in terms of storage requirements, packing requirements and transportation. By flying between rows of trees, UAV cameras can capture images that can be used for estimation of fruit yield. Due to surrounding objects and illumination differences, fruit count estimation is a challenging task with the added challenge of differentiating among overlapping fruits. Most traditional algorithms used the shape, texture and color of fruits in Computer Vision but these methods are fruit specific. Deep Learning algorithms perform better with complex environments consisting of noise elements. The steps as stated in Chen et al. (2017) are utilizing a crowd sourcing platform to obtain human generated labels on a database of fruit images. These labels are connected in the Vector based Scalar Vector Graphics (SVG) format. Users draw circles around the fruit present in the image so the classification is not pixel based. The final dataset consisted of 71 1280 x 960 images of oranges and 21 1920 x 1200 images of apples. Subsequently, a fully convolutional network (FCN) is used to output an estimated fruit count.

Finally, a linear regression model is trained to map the fruit counts in the previous layer to human generated labels. The work treats fruit counting as a counting problem rather than a simple pixel classification problem.

Obstacle Detection

The inclusion of robots and autonomous systems could be one of the most revolutionary steps in the effort to overcome the large difference between the demand and supply of quantities of food. These systems can assure maximized production of food supplies, with the efficient utilization of limited resources at disposal. Obstacle detection is one of the most important features of these systems which will allow them to be adapted into the field of smart and precision-based agriculture. To increase scalability and efficiency in the field of yield prediction and disease management, it is important to be able to use semi-automated and automated agents in the field, which can be of utility to detect diseases and be able to check for healthy specimens in the crop to predict yield. The development of such systems can be used to assist workers transfer payloads while also conducting different agricultural activities such as crop monitoring, weeding, etc.

Deep learning is being employed to detect obstacles using different classification and localization principles where the output of a normal image classification algorithm is modified to contain parameters of box boundaries over the different objects identified or detected in an image. These models can be trained to detect certain objects that are regarded as obstacles based on the application at hand. Using a combination of different obstacle detection and SLAM (Simultaneous localization and mapping) techniques, the autonomous systems can perform a variety of different tasks.

Such a study was conducted in Ferentinos (2018) where the authors use deep convolutional neural networks to detect obstacles and objects that can cause hindrance to productivity in different agricultural activities. Facilitated by the process of creating a heat map of the images in the dataset, providing an intensity-based representation to different objects in the image.

The authors augment their dataset by randomly sampling sub-areas of the extracted training samples, which can be said to important in applications where the image dataset is not of very big size. The approach in the cited work is based on the fine tuning of Alex net which is pre trained on ImageNet as discussed in the plant pheno typing application. The output of the original Alex net is modified to a two-dimensional vector, since in this case the requirement is of being able to detect if an image area contains the object or not. The fine tuning took fourteen thousand iterations in a batch size of hundred images. The learning rate in the fine-tuning state was less by a factor of 10 than the learning rate of the fully connected layers.

The network was robust enough to run on sample of images of any size, yielding a heat map of the image highlighting the obstacle. The researchers reduced the resolution of the sample images by a factor of two in order to reduce processing time. Post the training phase of the DCNN, the detection of the object is done based on the intensity of the heat map that was obtained as a result of the DCNN. This model reported an accuracy of 99.9% in row crops and 90.8% in grass mowing. These levels of accuracy are much better than normal human capabilities.

Such applications as explained before, can be regarded as game changers in the field of smart and precision-based agriculture.

Land Cover Classification

Information about the different aspects of land resources are of utmost importance for better land allocation and utilization to different tasks. It also helps to be able to study the different dynamic relationships of the ecosystem of an area, which can for one aspect, effect the agricultural produce in an area, it can also be used to identify the prominent diseases in the area, which can allow the farmers to be better prepared while managing diseases. Land use and land cover classification (LULC) based maps are used extensively in order to identify these relationships between different contributing factors of the land resource ecosystem, and the creation of these maps is done through remote sensing, using common image classification algorithms. Most approaches meant to analyze these maps are based on assumption of cloud free and mono-temporal samples. Since clouded image samples tend to be of little use for analysis purposes, most approaches tend to apply a cloud filtering as a pre-processing step. Plus taking in mono-temporal samples to study dynamic relationships in vegetation-based classes whose spectral reflectivity changes depending on different biological processes depending on the environmental conditions and vegetation type, might not be the most appropriate methodology. Since more processing power has become available along with abundance of sample data, the previous constraints acting as barrier to temporal feature-based analysis can be said to be lifted. This is said to be the perfect use case of the application of deep learning-based sequence models. The temporal analysis alleviates the requirement of cloud filtering step by considering the atmospheric perturbations as temporal noise, as it assumes that the sequential network would have learned the cloud filtering schemes solely from input data. The combination of different LSTM units in a RNN is said to be able to remember different context-based features which forms the basis of this application. A study to estimate the performance of RNNs with LSTM units was done in [20]. It involves the use of bi-directional RNNs in which the sequence of images is passed to create a final memory cell value which is considered the final state of the entire sequence, and is passed on to convolutional layers to classify the regions. This study concluded that the combination of LSTM with SVM give the best results for this problem when compared to other options.

CONCLUSION

In this study, the need for smart agriculture in the field of Management of Agricultural Diseases and Yield Prediction is explained. The importance of computer vision, and how deep learning affects the field is also discussed. Then concepts, techniques and algorithms of Deep Learning beneficial to smart agriculture are explained. These algorithms form the basis for the study of different applications of deep learning in smart agriculture. The applications discussed include Fruit Counting, Obstacle Detection, Land Cover Classification, Plant Phenotyping, Weed Identification and Plant Disease Detection.

These applications are of major impact in the fields of Agricultural Disease Management and Yield Prediction. Where applications like Fruit Counting, Plant Phenotyping, Weed Identification and Plant Disease Detection are of direct consequence in the fields, applications like Obstacle Detection and Land Cover Classification, though indirectly, heavily effect the supporting mechanisms required for these sub-fields in the area of smart agriculture.

As explained in detail, approaches in the field of Smart and Precision based agriculture, coupling traditional methods with deep learning are giving state of the art results. Hence, it would be fitting, to conclude, that these methods are the way forward, in order to bridge the gap between the millions who cannot afford proper meals around the world.

MOVING FORWARD

The authors recommend the readers to study deep learning in detail understanding the implementation of different famous frameworks like TensorFlow, Keras and PyTorch which make neural networks much more accessible. The authors also recommend going through other applications such as weather forecasting, animal behavior studies, event date estimation, soil moisture studies and leaf area index, which are also of prime importance to the topic at hand, resources to which have been provided in the additional reading section.

REFERENCES

Bah, M.D., Hafiane, A., & Canals, R. (2018). *Deep Learning with unsupervised data labeling for weeds detection on UAV images.* CoRR, abs/1805.12395

Bishop, C. (2006). *Pattern Recognition and Machine Learning*. New York: Springler-Verlag.

Chen, S.W., Shivakumar, S.S., Dcunha, S., Das, J., Okon, E., Qu, C., Taylor, C.J. & Kumar. (2017). Counting apples and oranges with deep learning: A data-driven approach. *IEEE Robotics and Automation Letters*, *2*(2), 781-788.

Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture*, *145*, 311–318. doi:10.1016/j.compag.2018.01.009

Food & Agricultural Organisation. (2009). *How to Feed the World in 2050*. Retrieved from http://www.fao.org/fileadmin/templates/wsfs/docs/expert_paper/How_to_Feed_the_World_in_2050.pdf

Gers, F., Schmidhuber, J., & Cummins, F. (1999). Learning to forget: continual prediction with LSTM. In *1999 Ninth International Conference on Artificial Neural Networks ICANN 99*. Edinburgh, UK: IET, . 10.1049/cp:19991218

Gerven, M., & Bohte, S. (Ed.). (2017). Artificial Neural Networks as Models of Neural Information Processing. Frontiers' Research Topics.

Goh, A. T. C. (1995). Back-propagation neural networks for modeling complex systems. *Artificial Intelligence in Engineering*, *9*(3), 143–151. doi:10.1016/0954-1810(94)00011-S

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

Han, S., Wang, Y., Yang, H., Dally, W. J., Kang, J., Mao, H., & Yao, S. (2017). ESE. *Proceedings of the 2017 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays - FPGA'17*. doi:10.1145/3020078.3021745

Jin, J., Fu, K., & Zhang, C. (2014). Traffic Sign Recognition With Hinge Loss Trained Convolutional Neural Networks. *IEEE Transactions on Intelligent Transportation Systems*, *15*(5), 1991–2000. doi:10.1109/TITS.2014.2308281

Kosa, J., Bennett, F., Andre, D. & Keane, M. (1996). *Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming*. Academic Press.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Communications of the ACM*, *60*(6), 84–90. doi:10.1145/3065386

Lipton, Z., Berkowitz, J., & Elkan, C. (2015). *A Critical Review of Recurrent Neural Networks for Sequence Learning.* Retrieved from https://arxiv.org/abs/1506.00019

Namin, T., Esmaeilzadeh, M., Najafi, M., Brown, T. B., & Borevitz, J. O. (2017). *Deep phenotyping: Deep learning for temporal phenotype/genotype classification*. . doi:10.1101/134205

Nielson, M. (2015). *Neural Networks and Deep Learning*. Determination Press.

Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Article. *Deep Face Recognition.*, *1*(3), 1–12.

Razavian, A. S., Azizpour, H., Sullivan, J., & Carlsson, S. (2014). CNN features off-the-shelf: an astounding baseline for recognition. *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 10.1109/CVPRW.2014.131

Rußwurm, M., & Körner, M. (2018). Multi-Temporal Land Cover Classification with Sequential Recurrent Encoders. *ISPRS International Journal of Geo-Information*, *7*(4), 129. doi:10.3390/ijgi7040129

Sak, H., & Senior, A. W. (2017). *Processing acoustic sequences using long short-term memory (LSTM) neural networks that include recurrent projection layers*. U.S. Patent No. 9,620,108.

Steen, K., Christiansen, P., Karstoft, H., & Jørgensen, R. (2016). Using Deep Learning to Challenge Safety Standard for Highly Autonomous Machines in Agriculture. *Journal of Imaging*, *2*(1), 6. doi:10.3390/jimaging2010006

United Nations. (2017). *World population projected to reach 9.8 billion in 2050, and 11.2 billion in 2100*. Retrieved from https://www.un.org/development/desa/en/news/population/world-population-prospects-2017.html

Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P. A. (2008). Extracting and composing robust features with denoising autoencoders. *Proc. 25th Int.Conf. on Machine Learning*, 1096–1103.

Wechsler, H. (Ed.). (1992). Neural Networks for Perception. Academic Press, Inc.

Xia, X. L., & Han, R. Y. (2017). Research on classifications of mobile networking device users based on BP neural networks. *2nd International Conference on Mechatronics and Information Technology (ICMIT 2017)*, 104–109.

Zeiler, M., & Fergus, R. (2014). *Visualizing and Understanding Convolutional Networks*. Retrieved form https://arxiv.org/pdf/1311.2901.pdf

ADDITIONAL READING

Biswas, S. K., Sinha, N., Purkayastha, B., & Marbaniang, L. (2014). Weather prediction by recurrent neural network dynamics. *International Journal of Intelligent Engineering Informatics*, *2*(2-3), 166–180. doi:10.1504/IJIEI.2014.066208

Chai, L. N., Qu, Y. H., Zhang, L. X., Liang, S. L., & Wang, J. D. (2012). Estimating time-series leaf area index based on recurrent nonlinear autoregressive neural networks with exogenous inputs. *International Journal of Remote Sensing*, *33*(18), 5712–5731. doi:10.1080/01431161.2012.671553

Christiansen, P., Hansen, M., Steen, K., Karstoft, H., & Jørgensen, R. (2015). Advanced sensor platform for human detection and protection in autonomous farming. In *Precision Agriculture'15* (pp. 1330–1334). Wageningen, the Netherlands: Wageningen Academic Publishers. doi:10.3920/978-90-8686-814-8_35

Gerven, M. (Ed.). (2017). *Artificial Neural Networks as Models of Neural Information Processing*. Front. Comput. Neurosci.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

Lu, H., Fu, X., Liu, C., Li, L. G., He, Y. X., & Li, N. W. (2017). Cultivated land information extraction in UAV imagery based on deep convolutional neural network and transfer learning. *Journal of Mountain Science*, *14*(4), 731–741. doi:10.100711629-016-3950-2

Lyu, H., Lu, H., & Mou, L. (2016). Learning a transferable change rule from a recurrent neural network for land cover change detection. *Remote Sensing*, *8*(6), 506. doi:10.3390/rs8060506

Mohri, M., Rostamizadeh, A., & Talwalker, A. (2018). *Foundations of Machine Learning*. MIT Press.

Salman, A. G., Kanigoro, B., & Heryadi, Y. (2015). Weather forecasting using deep learning techniques. *Advanced International Conference on Computer Science and Information Systems (ICACSIS)*, IEEE. 10.1109/ICACSIS.2015.7415154

Steen, K. A., Christiansen, P., Karstoft, H., & Jørgensen, R. N. (2016). Using deep learning to challenge safety standard for highly autonomous machines in agriculture. *Journal of Imaging*, *2*(1), 6. doi:10.3390/jimaging2010006

Zhu, N. Y., Liu, X., Liu, Z. Q., Hu, K., Wang, Y. K., Tan, J. L., ... Guo, Y. (2018). Deep learning for smart agriculture: Concepts, tools, applications, and opportunities. *International Journal of Agricultural and Biological Engineering*, *11*(4), 32–44. doi:10.25165/j.ijabe.20181104.4475

Chapter 5
Computer Vision for Green Plant Segmentation and Leaf Count

Praveen Kumar J.
National Institute of Technology Tiruchirappalli, India

Domnic S.
National Institute of Technology Tiruchirappalli, India

ABSTRACT

Image-based plant phenotyping plays an important role in productive and sustainable agriculture. It is used to record the plant traits such as chlorophyll fluorescence, plant growth, yield, leaf area, width and height of plants frequently and accurately. Among these plant traits, plant growth is an important trait to be analyzed that directly depends on leaf area and leaf count. Taking benign conditions of quick advancement in computer vision and image processing algorithms, many methods have been developed in recent days to find the leaf area and leaf count accurately. In this chapter, the recent techniques in image-based plant phenotyping and their limitations are discussed. Also, this chapter discusses a new plant segmentation method based on wavelet and leaf count methods based on Circular Hough Transform and deep learning model, which overcomes the drawbacks of recent methods. These methods are experimented with Computer Vision Problems in Plant Phenotyping (CVPPP) benchmark datasets.

INTRODUCTION

Agriculture plays an important role in the economy of the country. The increase of population in a country leads to demand in agricultural products and hence, there is a need to increase the productivity of agricultural products. So, the agriculture related industries and the researchers are involving in research with great efforts, to continue agriculture for a prolonged period without any break. The important component of production statistics is yield rates. Most of the sampling operations for yield prediction are carried out manually. This may lead to possible occurrence of manual error and sampling error. To improve the

DOI: 10.4018/978-1-5225-9632-5.ch005

productivity, these errors have to be reduced by making them as automation process. Computer vision techniques can be used to increase the productivity and reduce the production costs by automating the process and increasing the accuracy. This leads in arising of exciting problems in the field of computer vision. One of such problems is image-based plant phenotyping.

The Plant phenotyping refers to a quantitative description of anatomical, physiological, ontogenetical and biochemical properties of the plants. In recent times, several image-based methods for plant phenotyping are developed which are gaining more importance and, on par with growing commercial and scientific interest. The Plant phenotyping addresses the rural requirements without any limitation. One of the important rural requirements is to improve the crop yield, which requires an exceptional amount of research. The key plant phenotyping traits like plant yield, growth and development of plants can be characterized by total number of leaves in the plant and the leaf area (green plant region). The Plant image analysis can be used for predicting the crop yield. The Plant image analysis deals with plant measurements like anatomy, growth, surface, shape, etc. by analyzing the images of various plant organs such as root, leaves, etc.

Currently, the majority of the plant phenotyping systems depend on various custom-built or commercial solutions, ranging from the small-scale controlled environment to the field applications or automated large-scale greenhouses. The commercial solutions are more expensive and also require initial investment, which can be afforded by only few laboratories. Also, commercial solutions require appropriate analysis software and thus becoming vendor locked. Based on these limitations, many organizations embrace image-based plant phenotyping methods, which are localized for their environment, rather than using commercial solutions. Such image-based plant phenotyping methods are capable of addressing only specific phenotyping problems. Moreover, these image-based plant phenotyping methods cannot be easily implemented in different environments with various experimental settings, since many modifications in the design or in the complex image pipeline are necessary.

A vital amount of research have been performed in the field of plant phenotyping, which includes the research in leaf region segmentation, leaf counting, plants disease identification, and observing the development and growth of plant by analyzing the plant images. The image-based plant phenotyping is used to analyze the characteristics of plant growth and development, for estimating the crop yield. Plant growth depends on the total number of leaves and leaf area (Orlando et al., 2011), hence the leaf count measurement will be used for assessing the plant growth. The plant images may contain numerous leaves, branches, stems, and other objects in the background which meddle with the procedure. The leaf region must be isolated from the image, in order to count the leaves accurately. This chapter analyzes several methods existing in the field of image-based plant phenotyping and their limitations. Also, this chapter focuses on developing a new method which overcomes the limitations of existing methods.

Background

In recent times, plant phenotyping has been performed in controlled laboratory environment, green house, or in the field. Many academicians and researchers have contributed a lot to plant phenotyping. Huang and Lee (2010) proposed a method for automatic identification of plant species. Grand-Brochier et al., (2015) depicted the studies to extract the tree leaves from natural images by applying various segmentation methods. Tang, Liu, Zhao and Tao (2009) proposed an algorithm to segment leaves from complex background. Some of the segmentation systems use infrared or depth information (Chéné et al., 2012).

Yin, Liu, Chen and Kramer (2014) proposed a method to segment and track the Rosette leaves which matches the existing segmented leaf templates with hidden leaves. Dellen, Scharr and Torras (2015) proposed a graph-based method to track the leaves of tobacco plant. De Vylder et al., (2011) proposed a method based on active contour to segment and track Rosette arabidopsis leaves. Shen, Zhang and Chen (2007) proposed computer vision based automated system to count the soybean leaf aphids. Cerutti et al., (2011) proposed a parametric active polygon model. The limitations of these methods are (i) difficult to handle occlusions and (ii) need of large labeled datasets and prior training.

In recent days, numerous research works have been carried out to measure the plant phenotypic traits (De Vylder et al., 2012; Arbelaez et al., 2011; Ning et al., 2010; An et al., 2016; Scharr et al., 2016). The open source Rosette Tracker tool is used to evaluate the genotype effects (De Vylder., 2012). The gPb-owt-ucm is a segmentation method based on contour detection and spectral clustering (Arbelaez et al., 2011). Maximal Similarity Based Region Merging (MSRM) is an interactive segmentation method which uses region merging framework to fuse the super-pixel segmentation (Ning et al., 2010). The limitation of these methods is that these methods result in either under segmentation or over segmentation.

Giuffrida et.al., (2016) used log-polar representation and global descriptors (GLC) to estimate the number of leaves in a plant. An et al., (2016) showed the practical workflow for higher image processing throughput. The phenotypic traits such as rosette area, total leaf expansion and leaf length are extracted from input images using photogrammetry and imaging techniques. In this work, the images are captured in movable field platform and stationary growth chamber platform using several cameras placed at various angles which results in occurrence of shadows, color changes and distortion in the image. Hence, the authors have used image preprocessing techniques such as Image optical distortion correction, Image color correction and Orthophoto generation. Then, the Normalized Green–Red Difference Index is used for plant region segmentation from the rosette plant images. This segmented plant region is used for calculating the rosette area, total leaf expansion and leaf length. The limitation of this framework is that it was camera dependent (i.e.) not generalized for all brands of cameras. This framework needs corresponding preprocessing software for any brand of camera. The Normalized Green–Red Difference Index is used to segment the plant region in this framework. Since, the leaf region and mosses are green in color the Normalized Green–Red Difference Index may not differentiate the green plant region from the mosses. Therefore, it may result in over segmentation. Praveen Kumar and Domnic (2017) proposed a histogram-based quantization method (Hist_Quantize) to segment the plant region. Praveen Kumar and Domnic (2018) proposed a Graph-based quantization method (Graph-based) to segment the plant region. They have used various color models for plant segmentation.

In current decade, the Leaf Segmentation Challenge (LSC) and the Leaf Counting Challenge (LCC) have been conducted as part of the Computer Vision Problems in Plant Phenotyping (CVPPP) workshop, which has been held in conjunction with the several Image Processing and Computer Vision Conferences. Many methods have been presented in the challenges. Few methods have been successful in segmenting the leaf region or counting the plant leaves. These methods are tested on CVPPP benchmark datasets. Few methods based on color models (Scharr et al., 2016) and deep learning models are discussed in detail as follows:

IPK Gatersleben: Segmentation Via 3D Histograms

The IPK pipeline is based on unsupervised clustering and distance maps for segmenting the leaves. The workflow is summarized as following.

1. Supervised foreground/background segmentation using 3D histogram cubes, which determine the probability for a pixel colour in the given training set of images to the foreground or background; and
2. Unsupervised feature extraction such as detecting the leaf split point and leaf centre points for individual leaf segmentation by using skeleton, distance map, and the corresponding graph representation.

In this work, a direct look-up in the 3D histogram cubes is used rather than using several thresholds for one-dimensional colour component. The two 3D histogram cubes for plant region (foreground) and non-plant region (background) are accumulated using the training data. The input images are converted to the L*a*b colour space in order to improve the performance of the framework against illumination variability. The estimation of the entries which do not present in the training data is done by using an interpolation of the nearby values in the histogram cell. The segmented plant region may have noise region. The noise regions are suppressed by performing morphological operations and cluster-analysis of the segmented plant region. The output of this operation is given as input for the feature extraction phase to identify the optimal split points and leaf centre points of corresponding leaf segments.

For this framework, the leaves of Rosette plants are considered to be compact. The Rosette plant leaves may partly overlap and thus producing the partial leaf occlusion. Since, the leaf centre points appear as peaks in the corresponding Euclidean distance map, they are identified by a maximum search. The skeleton image is calculated as the next step. The split points at the thinnest connection points are identified to resolve overlapping in the Rosette leaves. The values in the Euclidean distance map are mapped to the skeleton image. The output image of this mapping is used to form a skeleton graph. The skeleton branch-points, skeleton end-points, and leaf centre points are represented as nodes in the skeleton graph. The edges in the skeleton graph are created if there is a connection between corresponding image points by a skeleton line. A list of minimal distances and the positions of every particular edge are saved as an edge-attribute. This list is used to find the positions of the leaf split points accurately. Then, all the nodes and edges of the graph that connects the two leaf centre points are traversed and the position with the minimal Euclidean distance map values is determined to detect the split points between two leaf centre points. This process of detecting the split points is repeated, when there is a connection between the two leaves which have to be separated.

The two coordinates on the border of the plant leaf are calculated to identify the split line belonging to the corresponding minimal EDM point. The nearby background pixel (first point) is searched, and also the nearby background pixel at the opposite position in relation to the split point (second point) is identified. The leaf border is represented by the connection line during the segmentation of occluded leaves. In the final step, the individual leaves are labeled by using region growing algorithm.

Nottingham: Segmentation With SLIC Superpixels

This method uses superpixel for segmenting the leaves from the plant images. It does not require any training. The training dataset (CVPPP benchmark datasets) has been used for tuning segmentation parameters only. The steps used in this method are given as follows:

1. Superpixel over-segmentation in L*a*b colour space using Simple Linear Iterative Clustering (SLIC);

2. Segment the Rosette plant region (foreground) using simple seeded region growing method in the superpixel space;
3. Calculate the distance map on segmented plant region;
4. Find the superpixels whose centroid is in the local maxima of the distance map of segmented plant region and match the individual leaf seed; and
5. Segment the individual leaf by applying watershed transform with the extracted leaf seeds.

The steps (1) and (2) are used in segmenting the Rosette plant region and the steps (3), (4) and (5) are used in segmenting the individual leaves.

Initially, the input RGB image is converted to L*a*b colour space for enhancing the difference between plant region and non-plant region. Then, the superpixels are calculated. A fixed number of SLIC superpixels are calculated over the input image. Experimentally, it was found that 2000 pixels provide better coverage of the leaf region. The mean value of the 'a' channel (green colour of the image is characterized by 'a' channel of L*a*b colour space) is extracted for the entire superpixel. A superpixel neighborhood graph (Region Adjacency Map) is created from the resulting superpixels. In order to segment the plant region, a simple region growing approach is applied using the mean value of every superpixel in channel 'a'. The initial seed for region growing process is selected based on the lowest mean value for superpixel in channel 'a' of L*a*b colour space. However, the process of plant region segmentation can be made even faster and more accurate by applying simple thresholding of mean value for each superpixel. The thresholds applied in this approach are -25 and -15 for A1 and A2 datasets of CVPPP benchmark datasets respectively.

After segmenting the plant region from the input plant images, the superpixels that do not represent the plant region are removed. Then, the strong edges which can be identified using the Canny edge detector are removed. A distance map is computed and the centroid is calculated for all the superpixels. The superpixels which exist in the centre of the leaves are identified by applying the local maxima filter. A superpixel is chosen as a seed when it lies in the most central part of the leaf compared to its neighboring superpixels within a radius. This is implemented by looking up the distance map; considering the centroid value of superpixel and filtering the superpixel which do not have maximum value among the neighbors. Finally, the individual leaf region is segmented by applying the watershed transform using the obtained initial seeds over the image space.

The advantage of this method is that it is a fast method. Also, this method does not require any training. The training datasets are used for tuning the parameters only. The accuracy of plant region segmentation and individual leaf region segmentation can be improved when this method is tuned on per image basis. The following are the parameters which can be tuned to improve the accuracy of plant region segmentation and individual leaf region segmentation

1. Number of superpixels.
2. Foreground extractor (region growing or threshold)
3. Compactness of superpixels.
4. Parameters of the Canny edge detector
5. Colour space for SLIC, canny edge detector and foreground extractor.

In order to increase the accuracy of this approach, these parameters were tuned in a per-dataset basis. If required, these parameters can also be manually tuned on a per-image basis.

MSU: Leaf Segmentation With Chamfer Matching

The MSU approach is an extension work of multi-leaf alignment and tracking framework. This is the approach which uses the plant fluorescence videos for plant region segmentation because these videos will have clean background. This method adopts advanced background segmentation process. This approach is based on Chamfer Matching (CM) technique in which the instance of an object in an image is aligned with a given template. Since, there are huge variations in size and shape of leaves in the plant, it is impossible to find the individual leaves in the plant with the single template. Hence, a large number of templates are created with different shapes, orientations and scales. For training this approach, H number of leaves with various aspect ratios is chosen from images of the training set.

Specifically, H leaves with representative shapes (e.g. different aspect ratios) are selected from H images of the training set. Each shape of the leaves is scaled to different sizes (S) which is rotated to various orientations (R). Thus, a set of $H \times S \times R$ leaf templates with labelled tip locations are created. These templates are used for segmenting the leaf region input plant images. It is very critical to find the reliable Chamfer Matching results using accurate edge map and plant region segmentation. Hence, the input images are converted to L*a*b colour space. Then, a threshold τ (chosen experimentally for each dataset in CVPPP benchmark datasets: 40 for A1 dataset, and 30 for A2 and A3 datasets) is applied to 'a' channel of L*a*b colour space to estimate the foreground mask or plant region. This segmented plant region is subject to refinement by standard morphological operations. An edge map is generated by applying the Sobel edge operator within the segmented foreground or plant region. The dataset A3 has plant images with more overlapping leaves and shadow region. The leaf boundaries are not clearly visible due to the presence of shadow region. Hence, an additional edge map is required for segmenting the individual leaf region. This is generated by applying the Sobel edge operator over the resulting image obtained by the difference of 'a' and 'b' channels of L*a*b colour space. The standard morphological operations are applied to remove the noise region (small edges) and leaf veins (lines).

A mask and the edge map are cropped from the input image. For each template, identify one location in the edge map with the minimum CM distance. Repeat this process of identifying the locations having minimum CM distance for each template. Now, an over complete set of leaf candidates is generated. For each leaf candidate, the CM score, its overlap with plant region, and the difference in angle are computed. Next, the subset of leaf candidates has to be selected as the result of plant region segmentation. First, the leaf candidates with high CM scores, a large difference in angle, and a small overlap with the foreground or plant region are deleted. Second, an optimal set of leaf candidates is selected by using optimization process which optimizes the smaller number of leaf candidates with smaller differences in leaf angle and smaller CM distances to cover the segmented plant region as much as possible. Third, the redundant leaf candidates have to be deleted. For this, all the leaf candidates are considered for initialization and the gradient descent is applied iteratively. This is used to identify the final set of non-redundant leaf candidates. It is impossible to perfectly match all the edges in images with the finite number of templates. Hence, the multi-leaf tracking procedure is applied on each template for transformations like rotation, translation, and scaling, such that the best possible match with the edge map is obtained.

This can be performed by minimizing the summation of the terms such as average CM score, average angle difference, and the difference between the synthesized mask of the test image mask and all candidates. The result of the leaf alignment provides initialization parameters for the transformation; and gradient descent is applied to update the parameters. The leaf candidate is removed when it becomes smaller than a threshold. After the optimization process, the leaf candidates match more accurately with the edge map. This overcomes the limitations of small number of final set of leaf templates. Finally, the foreground mask and the tracking result are used to create the label image such that all the plant region pixels have labels. To generate the templates, only one leaf from each of H training images is used. The similar kind of pre-processing procedure and segmentation procedure are carried out independently for every image in the training and testing dataset.

Wageningen: Leaf Segmentation With Watersheds

This method comprises two steps:

1. Plant segmentation.
2. Separate leaf segmentation.

Plant region is segmented from the background using neural network based supervised classification. Since the CVPPP benchmark datasets (A1, A2 and A3) are different in nature, a different classifier and post-processing steps are used. The ground truth images are used as the mask to segment plant region and background region. For training this approach, three thousand pixels for each image in every class are randomly selected. When the plant is smaller than three thousand pixels, all the pixels in plant region are used.

Two texture features and four colour features are used for each pixel to separate the plant region and background region. The four-color features used for classification are red (R), green (G), and blue (B) pixel values and the excessive green (2G–R–B) value that highlights the green pixels. The two texture features used in this approach are the pixel values of the gradient magnitude filtered green channel and the pixel values of the variance filtered green channel. These two features indicate the rough parts and edges in the image. To get the better results, a large range of nonlinear and linear classifiers have been tested for CVPPP benchmark datasets using the feed-forward MLP neural network with single hidden layer of 10U.

The plant masks (i.e. foreground–background segmentation) are created by applying the morphological operations on the binary image obtained after plant classification. The morphological operations used for A1 and A2datasets include erosion and the propagation. This process mainly removes the small blobs due to the presence of mosses. For A3 dataset, this process removes all blobs in the image except the large one. The watershed transform is highly sensitive to presence of mosses and spaces between stem and leaves. Additional morphological operations, colour transformation, and shape and spatial filtering are required in order to remove the occurrence of mosses in A2 and A3 datasets and also to give emphasis on spaces between stems and leaves.

For A2 dataset, all the portions of segmented foreground region that are far away from centre of gravity of the foreground mask than 1.5 times estimated radius of the foreground mask, are filtered out. The radius r is estimated using the formula $r = (A/\pi)^{1/2}$ where A is the foreground mask area. Next, the Y plane (which denotes the luminance) of the image of YUV colour space is thresholded with a threshold which is optimized based on the training dataset (th = 85). For A3 dataset, there is presence of huge moss areas along with the foreground segmentation mask. In order to remove these noisy region, first calculate the compactness (C) of the foreground mask by the formula $C = L^2/(4\pi A)$, where L is the contour length of the foreground mask. When, the compactness $C > 20$ it indicates the occurrence of a large moss area along with the segmented plant region. Next, the X plane (which denotes the chromatic information) of the image of XYZ colour space is thresholded with a threshold value of 55), and those pixels which are lesser than the threshold value are filtered out. Thus, the presence of mosses (slightly different colour than plant region) in the segmented plant region is removed. Then, in order to give emphasis to the spaces between the stems and the leaves, the foreground masks are thresholded with the threshold value for the Y plane of YUV colour-transformed foreground image as explained for A2 dataset.

The second step of this approach is to perform the separate leaf segmentation. It is achieved by using the watershed transform applied on the Euclidean distance map of the segmented plant region image obtained as the result of first step of this approach. Initially, the watershed transform is applied without using the threshold between the basins. After that, the basins are merged successively if they are detached by a watershed that is lesser than the given threshold value. This threshold value is tuned based on the training dataset in order to give the best result. The threshold value is set to 30, 58, and 70 for A1, A2, and A3 datasets respectively.

Deep Learning Models for Plant Segmentation and Leaf Counting

In the recent times, various deep learning architectures are used in the field of plant phenotyping. The usage of deep learning dramatically increases the accuracy of the system. The deep plant phenomics (DPP) approach (Ubbens & Stavness, 2017) proposed a method which is addressing the problem of counting directly without both plant segmentation and instance segmentation. The architecture as well as input dimensions are customized to achieve state-of-the-art accuracy on different subsets of the LCC-2015 dataset. Aich and Stavness (2017) have used the Deep Convolutional and Deconvolutional Networks (DCDN) to count the Rosette plant leaves. In this method, the Segnet architecture is used to segment the plant region. Giuffrida et al. (2018) have used Resnet50 architecture to count the Rosette plant leaves from the multi-modal 2D images.

MAIN FOCUS OF THE CHAPTER

Issues of the Existing Methods

There are some issues existing in these methods. In ‘MSU: leaf segmentation with Chamfer matching’ method, the accuracy of leaf region segmentation depends on the total number of templates generated. In this method, if the input images have leaves with large variations in shape and size, there is a need to create more templates. Hence, the template creation becomes a complex task. And also, this method has the major drawback of less hourly throughput. In ‘Wageningen: leaf segmentation with watersheds’ method, the system has to be initially trained. The accuracy of the system depends on initial training. In this method, the ground truth images are used to create the plant mask. Since, this method relies on ground truth images, this method may not be used for cross platform. In ‘Nottingham: segmentation with SLIC superpixels’ method, the plant region segmentation accuracy depends directly on the selection of superpixels. Moreover, the ‘Wageningen: leaf segmentation with watersheds’ and ‘Nottingham: segmentation with SLIC superpixels’ methods did not equally perform for the images in the CVPPP benchmark datasets and thus, the robustness of the system is not maintained. The drawback of the DPP approach is that it is not known if the approach would generalize and if a single DPP network would provide consistent results across all datasets. The DCDN method depends on ground truth images. In this method, the ground truth images along with the segmented images have been used to count the leaves.

Based on the existing works, it is observed that the accuracy of the system is based on the preprocessing software for image enhancement, and template creation or training datasets for plant region segmentation. Also, dependency on ground truth images is also a drawback. To overcome the limitations of the existing systems, a new method is presented in this chapter which is based on wavelet for plant region segmentation. Also, the CHT and transfer learning approach are proposed for counting the leaves.

DATASET DESCRIPTION

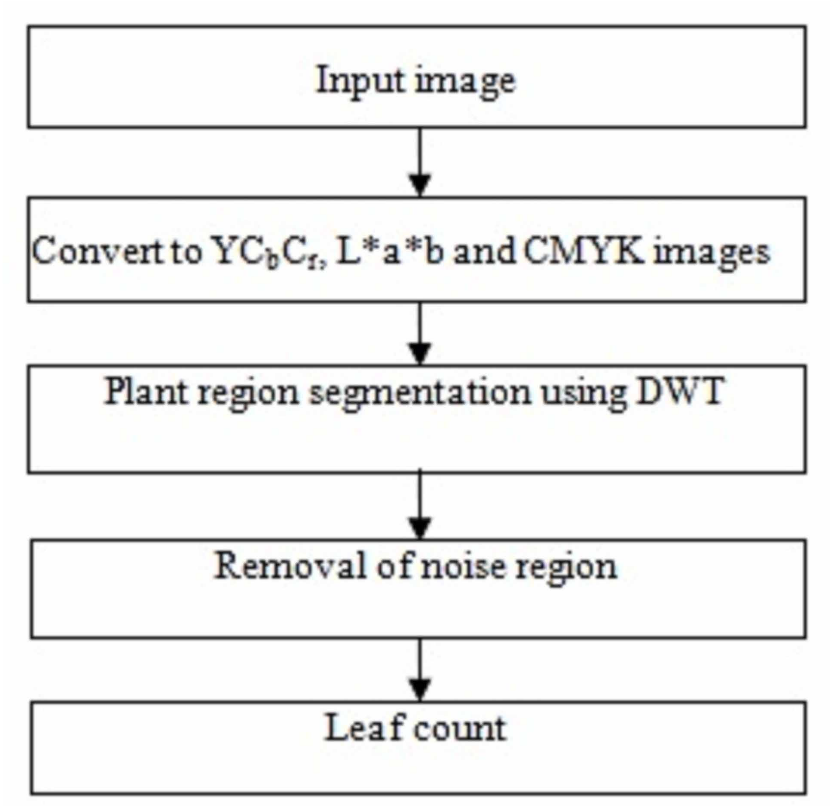
The datasets (A1, A2, A3) used for evaluating the proposed method are the CVPPP benchmark datasets (Scharr et al., 2014). The CVPPP benchmark datasets also consists of images of tray plants. The A1 and A2 datasets comprises top-view and time lapsed images of large number of Arabidopsis plants that are arranged in trays. The images of the plants are captured only in daytime for every 6 hour during the period of three weeks for A1 dataset, and for every 20min during the period of 7weeks for A2 dataset. These images are captured with a 7 MP Canon Power-Shot SD1000 camera. The raw images (pixel resolution of~0.167mm, 3108×2324 pixels) captured were saved as uncompressed (TIFF) files. Then, they were encoded by PNG file format lossless compression standard. The A3 dataset consist of tobacco plant images. The images in A1 dataset include the complexity of changing and complex background. The A2 dataset includes images with simpler scene. The images of A3 dataset have high image resolution. The A1 dataset consists of 128 images of 500×530 pixels resolution, A2 dataset consists of 31 images of 530×565 pixels resolution, and A3 dataset consists of 27 images with 2448×2048 pixels resolution. The CVPPP benchmark datasets include various plant images at different growth stages. The datasets A1, A2, and A3 include images of group of plants and isolated plants. These images have different scene complexity, resolution and fidelity. In terms of image analysis, the datasets include many challenges because of existence of many objects other than plants, and different scene complexities. The various complexities include light reflection due to presence of water in the tray, overlapping of leaves (i.e. leaf occlusions), varying shapes and sizes in different time because of nasty movements of leaves. In A3 dataset, the plant images contain nasty movement changes and varying leaf shapes. The leaf shapes are different because of various treatments. One of the treatments is growing the plants in different illumination condition. Under high illumination, the plants grow more compactly and also the leaves become partly wrinkled and occluded. Under low illumination, the leaves become round and large.

The A1 dataset includes the complexities such as changing background, slightly out of focus scene, external objects like tape or markers, occurrence of moss on the soil, yellowish dry soil. The A2 dataset includes the complexities such as various phenotype mutants, and different leaf size and shape. The complexities included in A3 dataset are high resolution, shadows, self-occlusion, leaf colour variation, and leaf hairs.

SOLUTIONS AND RECOMMENDATIONS

To overcome the drawbacks of the existing system, a new method is proposed to segment the plant region and to count the number of leaves in the plant. In the proposed method, initially the plant region is segmented from the background by applying the Discrete Wavelet Transform. Then, the number of leaves in the plant is counted using the Circular Hough Transform (CHT) and deep learning model (transfer learning approach). Figure 1 shows the steps in the proposed method.

Figure 1. Process of plant region segmentation and leaf count



Plant Region Segmentation

In the proposed method, the Discrete Wavelet Transform (Choudhary & Parmar, 2016) is used for segmenting the plant region to increase the robustness of the system. The Discrete Wavelet Transform is a technique which gathers the information about the location and frequency. Discrete Wavelet Transform uses wavelet filters to transform the image. There are different types of filters which can be used in transformation. They are

1. Haar Wavelet
2. Daubechies Wavelet
3. Bi-Orthogonal Filters,
4. Symlet wavelets
5. Coiflets wavelets
6. Meyer wavelets
7. Morlet wavelets etc.

The Discrete Wavelet Transform is chosen for segmenting the plant region because it has the capability to decompose an image at various levels of resolution and can be processed sequentially from low to high resolution by using wavelet decomposition. This is achieved because wavelets are localized in both frequency (scale) domain and time (space) domain. So, it is very easy to get to local features in an image. Moreover, the main advantage of using wavelet is that it can support multi resolution. By decomposing the image using wavelet, the sizes of the window differ and allow analyzing the image at various resolution levels. By using a wavelet transform, an image can be decomposed based on local frequency into various sub images. The wavelet shows an image as a sum of wavelet functions with various scales and location. There involves a pair of waveforms when an image is decomposed into wavelets. The high frequency known as wavelet function represents the detailed parts of an image. The low frequency known as scaling function represents the smooth parts of an image. The principle behind the decomposition of the wavelet is to transform the input image into many components with only one low resolution component known as approximation and the other components are known as details. Figure 2 shows the 1-level wavelet decomposition. To get finer details, the wavelet has to be decomposed to higher level. Figure 3 shows the 2-level wavelet decomposition and Figure 4 shows the multi-level wavelet decomposition.

In order to get the approximation component of an image, apply the bi-orthogonal low-pass wavelet in both horizontal and vertical directions and is trailed by a sub-sampling of every image using a factor of two for every dimension. To get the detail components of an image, apply the high-pass filter in one direction and a low-pass in the other direction or apply high-pass in both directions. The higher level of wavelet decomposition can be obtained by repeating the similar operations over the approximation component obtained at the previous level of wavelet decomposition.

The input images are considered to be MxN matrices where M represents the rows of the matrix and N represents the columns of the matrix. At each level of wavelet decomposition the horizontal data is filtered, and then the approximation component and detail components created from this are filtered on the columns. At each level wavelet decomposition, four sub images are formed, the approximation, the horizontal detail, the vertical detail, and the diagonal detail. The next level of wavelet decomposition is obtained by the decomposing of approximation component.

In the proposed method, Haar wavelet is used for plant region segmentation because it is the simplest and shortest basis. Also, it gives the satisfactory localization of characteristics of the image. The compact support property of the Haar wavelet enables it to decompose the image to have good time localization.

Figure 2. A 1-level wavelet decomposition

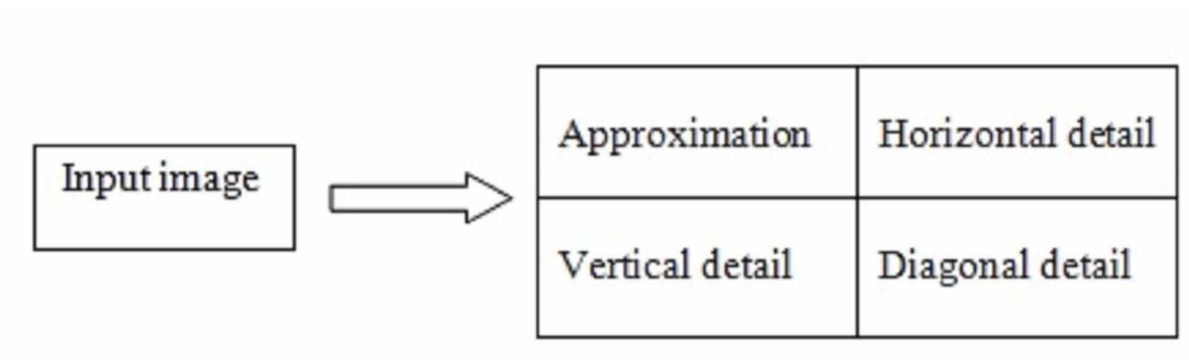


Figure 3. A 2-level wavelet decomposition

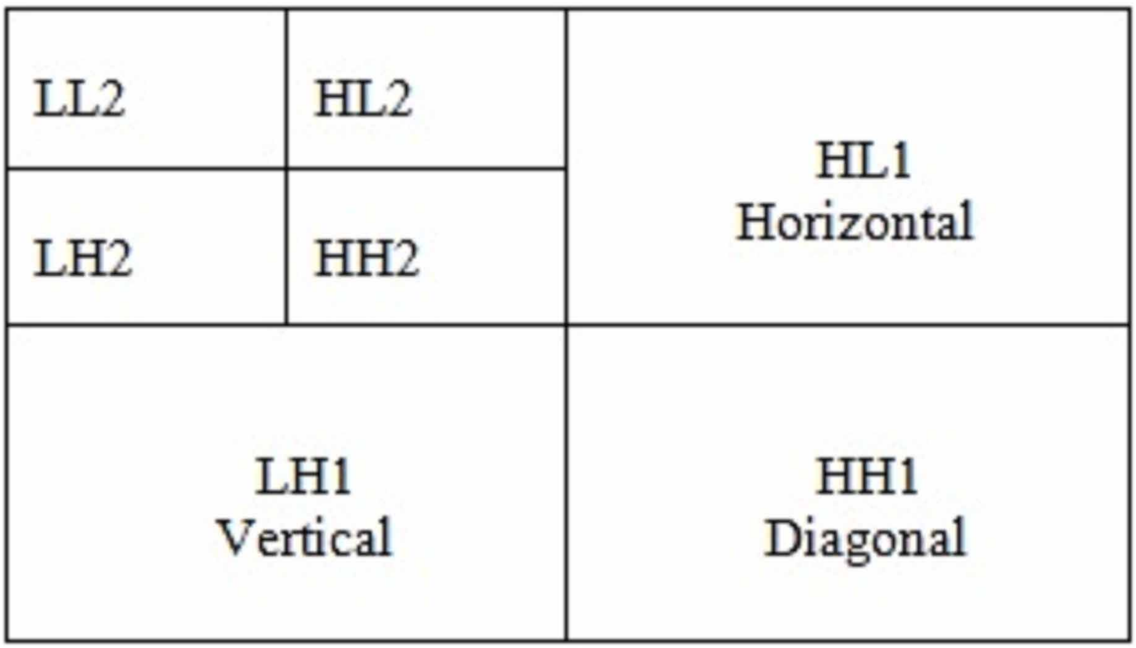
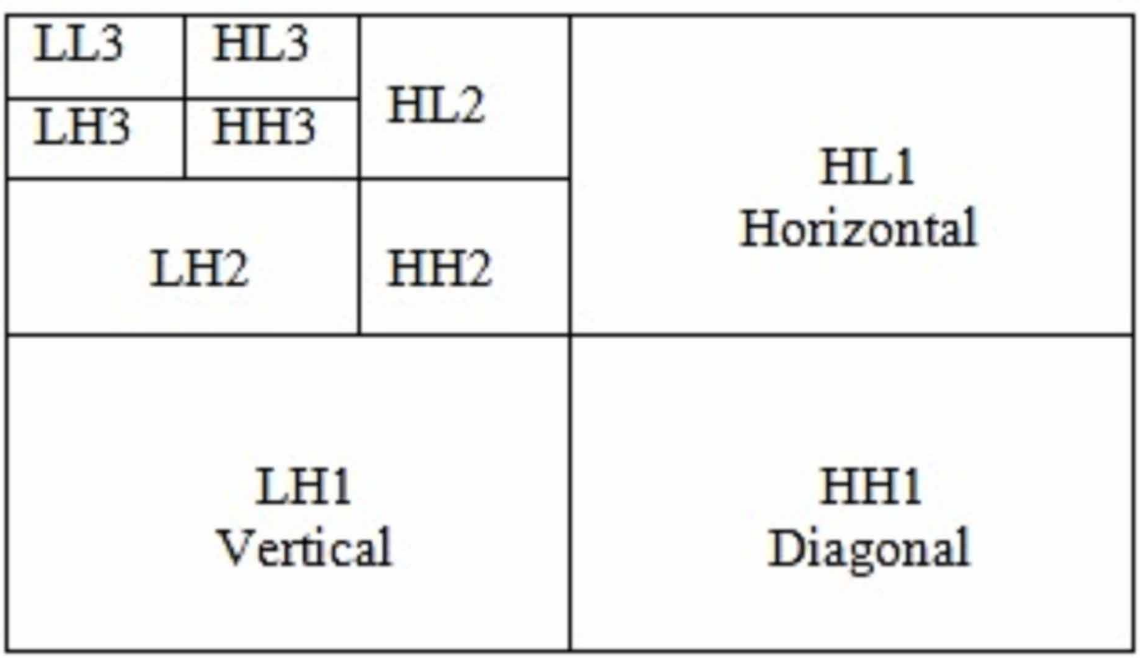


Figure 4. Multi-level wavelet decomposition



Explicitly, it represents that the Haar coefficients are successful in locating jump discontinuities and also represents the images efficiently with a small support. Haar Wavelet and Daubechies wavelet are also known as Maxflat wavelets because the frequency responses of both the wavelets have maximum flatness at Zero and R frequencies. This is a very advantageous property in some image processing applications.

In the proposed method, first the input image is converted to YC_bC_r color space. The Y component of YC_bC_r color space indicates the luminance value, the C_b component indicates the difference of blue and luminance component and the C_r component indicates the difference of red and luminance component. In C_b plane the pixel values show much difference between the shadow region and the visible region present in the green plant. But there is only small variation between the shadow region and the visible region of the green plant in C_r plane and therefore, leads to increase in the segmentation accuracy. So, in the proposed work the C_r plane is chosen for segmenting green plant region.

The C_r plane is given as input to the Haar Wavelet. A 2-level decomposition is performed (as described above) such that the green plant regions are reconstructed in grey color whereas the background region are reconstructed as white color. The reconstructed grey color plant region is used as mask to segment the green plant region from the input RGB image.

The segmented plant region may have noisy region such presence of light reflections, mosses and yellow soil. These noisy regions decrease the accuracy of plant region segmentation. To remove such noisy region, the segmented plant region is converted to L*a*b and CMYK colour spaces. Then, the thresholding technique is applied over these L*a*b and CMYK colour spaces. The *a plane and L plane of L*a*b colour space are used to get rid of the mosses. C and Y planes of CMYK colour space are used to get rid of other noisy regions.

The plant region that is segmented as a result of this phase is used for counting the number of leaves. The following are steps that describe the proposed plant region segmentation method using Discrete Wavelet Transform.

Step 1: Convert the input RGB image to YC_bC_r color space
Step 2: Apply 2-level Discrete Wavelet transform (Haar Wavelet) over the C_r plane
Step 3: Segment the plant region from the background using the reconstructed image that is obtained as a result of Step 2.
Step 4: Remove the noise region from the segmented plant region to improve the segmentation accuracy.

Leaf Count

In leaf count phase, the number of leaves is counted from the segmented plant region using CHT and deep learning model.

CHT Based Leaf Count

Since, the Arabidopsis plant leaves are round in nature, the number of leaves in Arabidopsis plant are counted by using Circular Hough Transform (CHT) (Illingworth & Kittler.,1987; Praveen Kumar &Domnic., 2018). Using CHT, the circular patterns in the image are identified. The curves that correspond to the collinear points intersect at the common point. The intersection point gives the properties of the lines that cross through these collinear points. This denotes that finding the concurrent curves involved in obtaining the collinear points.

Hence, the Circular Hough Transform can be considered as the modified Hough Transform. The CHT transforms the feature points in image space to the accumulated votes in parameter space. Then, the votes are collected in an accumulator array (3D_Hough_Array (h_0, k_0, r)) for each feature point and all the combinations of parameter. The array elements which have more number of votes indicate the presence of circle. There are two steps involved in CHT. They are (i) identifying the center of circle (c) and (ii) identifying the radius of circle (r).

The center of the circle can be obtained by the constraint that the normal vectors to the boundary of circle intersect at the center of the circle. The edge detection operators are used to find the normal vectors ($v_1, v_2, \ldots, v_n$). In the second step, the histogram technique for $d = (x - h_0)^2 + (y - k_0)^2$, where ($h_0$, k_0) is the center (c) of the circle (obtained in the first step of CHT), is used to find the radius of the circle. The radius of the circle is determined by identifying the highest peak in histogram.

The segmented plant region obtained from the plant region segmentation phase is given as input to count the leaves. The CHT is applied over this segmented plant region. The separate leaves are recognized and counted based on the circular region of the leaves. The number of leaves in the plant can be counted by counting the number of center of the circle.

Deep Learning Based Leaf Count

In Deep Convolutional Neural Network (DCNN) based leaf counting phase, the segmented plant region is given as input to count the number of leaves. The DCNN model used in the proposed leaf count model is VGG16 (Simonyan and Zisserman., 2014) which is trained on ImageNet dataset. The architecture of the VGG16 used in the proposed method is shown in Figure 5. In VGG16, there are twelve convolutional layers, five of which are followed by max-pooling layers, and two fully-connected layers. The principles of convolutional layer, pooling layer and fully-connected layer are given below.

Convolutional layer: The convolutional layer identifies various local patterns in each local region of the segmented plant image and creates several feature maps by using convolutional filters. This can be represented as

$$\left(e_k\right)_{ij} = \left(W_k \otimes g\right)_{ij} + b_k \tag{1}$$

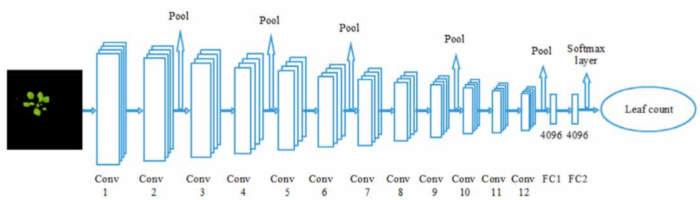
where $(e_k)_{ij}$ represents the (i,j) element of the k^{th} output feature map, W_k represents the k-th filter, g denotes the input feature maps, b_k represents the k^{th} bias and the symbol $\otimes$ denotes 2-D spatial convolution operation.

Pooling layer: The pooling layer is used to down-sample the feature maps created by the convolutional layer. Then, a single output from local regions of convolution feature maps is produced.

Fully-connected layer: The fully-connected layer integrates the preceding layers' outputs and generates the final feature representations for classification or regression. The activation function is a *sigmoid* or *tanh* function. The output is computed by the following equation.

$$o_k = \sum_l W_{kl} g_l + b_k \tag{2}$$

Figure 5. Fine-tuned VGG16 architecture



where o_k denotes the k^{th} output neuron, W_{kl} represents the weight connecting g_l and b_k, g_l denotes the l^{th} input neuron and b_k represents the k^{th} bias. To train the VGG16, Stochastic Gradient Descent is used with the weight decay value of 0.0005, and the momentum value of 0.9. It is given by

$$v_{i+1} = 0.9 \times v_i - 0.0005 \times \in \times w_i - \in \times \left\langle \frac{\partial L}{\partial f} \middle| w_i \right\rangle_{D_i} \tag{3}$$

$$w_{i+1} = w_i + v_{i+1}$$

where v represents the momentum variable, $\in$ is the learning rate and

$$\left\langle \frac{\partial L}{\partial f} \middle| w_i \right\rangle_{D_i}$$

is the mean of derivatives of batch D_i

In the proposed work, the VGG16 network pre-trained on ImageNet dataset is used to count the plant leaves. This trained VGG16 is fine tuned to find the leaf count from the segmented plant region. Since, the number of classes used in the pre-trained VGG16 model differs from that of the model used in leaf counting, the changes are made in the last two layers of the trained VGG16 model. Once, the VGG16 model is fine tuned, the output of last Fully Connected layer provides the features with dimensionality 4096. These features are used to find the leaf count.

The authors of the existing works in plant phenotyping have used various metrics to assess the performance of their works. They are

1. Foreground–Background Dice (FBD%): Used to find the delineation of segmented plant region with respect to the ground truth image,
2. Difference in Count (DiC): Used to calculate the difference between the obtained number of leaves in the plant image using the proposed counting method and actual number of leaves in the ground truth,
3. Absolute value of DiC (|DiC|),
4. Dice score (Dice%), as given in Equation (4), is used to determine the spatial overlap between the segmented plant region and the ground truth image,

5. Precision, as given in Equation (5), is used to find the portion of segmented plant region pixels that matches with the pixels of the ground truth image,
6. Recall, as given in Equation (6), is used to find the portion of pixels of the ground truth image present in the segmented plant region and
7. Jaccard, as given in Equation (7), used to find the spatial overlap between the segmented plant region and the ground truth image.

$$\text{Dice } (\%) = ((2 * TP) / ((2 * TP) + FP + FN)) * 100 \tag{4}$$

$$\text{Precision } (\%) = (TP / (TP + FP)) * 100 \tag{5}$$

$$\text{Recall } (\%) = (TP / (TP + FN)) * 100 \tag{6}$$

$$\text{Jaccard } (\%) = (TP / (TP + FP + FN)) * 100 \tag{7}$$

where the True Positive (TP) represents the number of pixels in the plant region that are correctly identified, the False Negative (FN) represents the number of pixels in the plant region that are not identified and the False Positive (FP) represents the number of pixels in the plant region that are falsely identified.

The authors in the paper (Scharr et al., 2016) used the metric FBD% to measure the accuracy of segmented plant region and the metric DiC to find the accuracy of counting the number of leaves in the plant image. The metrics such as Dice, Precision, Recall and Jaccard are used by the authors in their research works (De Vylder et al., 2012; Arbelaez et al., 2011) to measure the accuracy of plant region segmentation. Since, various metrics have been used by the authors of different research works to measure the accuracy of plant region segmentation, all these metrics can be used to evaluate the performance of the proposed method.

The proposed method is implemented in Matlab (release 2016), on a system with 4 GB memory, Intel i3 processor 2.66 GHz speed and, running on 64 bit windows operating system. The proposed method is tested over the CVPPP benchmark datasets (A1, A2, A3 and the whole tray plant images) under many challenging situations such as presence of yellow soil, reflection of light due to occurrence of water in the tray and moss growth. On an average, the time taken for segmenting the plant region and counting the number of leaves of each image in A1dataset and A2 dataset takes approximately 2 seconds and approximately 10 seconds for each image in A3 dataset. So, the hourly throughput of the proposed method for segmenting the plant region and counting the number of leaves is approximately 1800 plants per hour for A1and A2 datasets and approximately 360 plants per hour for A3 dataset. Figure 6, Figure 7 and Figure 8 show the experimental results on plant images from A1, A2 and A3 datasets.

The proposed method can be used in several plant phenotyping platforms in many environments with minimal modification in the threshold values of the color planes used in the noise removal phase. The proposed plant region segmentation method is compared with the state-of-the-art segmentation methods in plant phenotyping (Arbelaez et al., 2011; Kim et al., 2011; Zhang et al., 2013). Figure 9 shows the comparison of plant region segmentation results of the proposed method with these algorithms. The plant segmentation performance of the proposed method is better than the relevant recent methods and is shown in Table 1. It is observed from Table 1 that the proposed method performs better than the Nottingham and Wageningen methods, in terms of plant segmentation accuracy. The proposed method

Figure 6. Plant region segmentation and counting the number of leaves in plant images from A1dataset. (A) Original image, (B) Segmented plant region, (C) Leaf count.(Original leaf count at right top and Identified leaf count at right bottom).

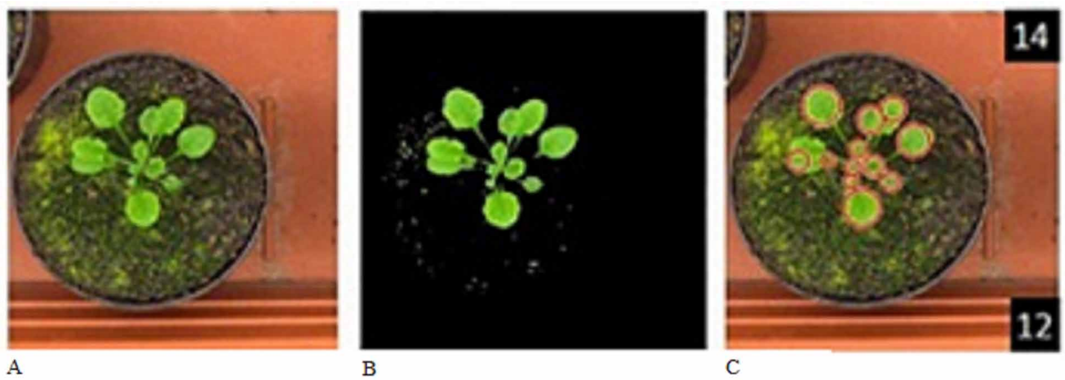


Figure 7. Plant region segmentation and counting the number of leaves in plant images from A2 dataset. (A) Original image, (B) Segmented plant region, (C) Leaf count.(Original leaf count at right top and Identified leaf count at right bottom).

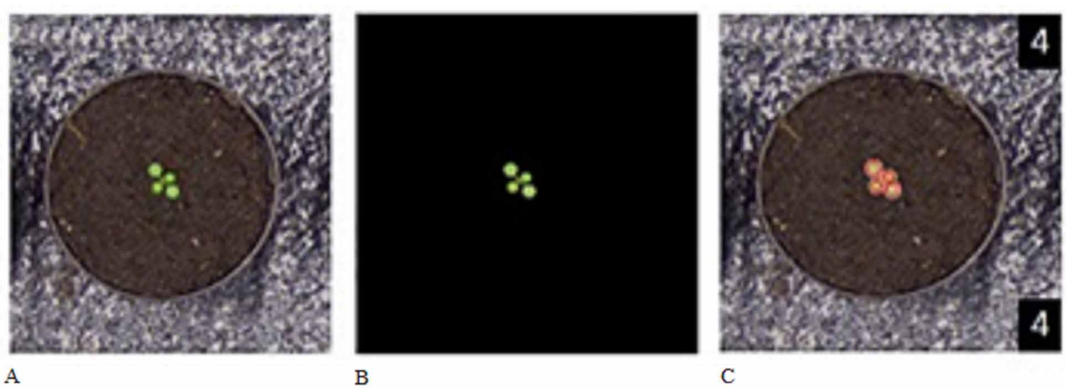


Figure 8. Plant region segmentation and counting the number of leaves in plant images from A3 dataset. (A) Original image, (B) Segmented plant region, (C) Leaf count.(Original leaf count at right top and Identified leaf count at right bottom).

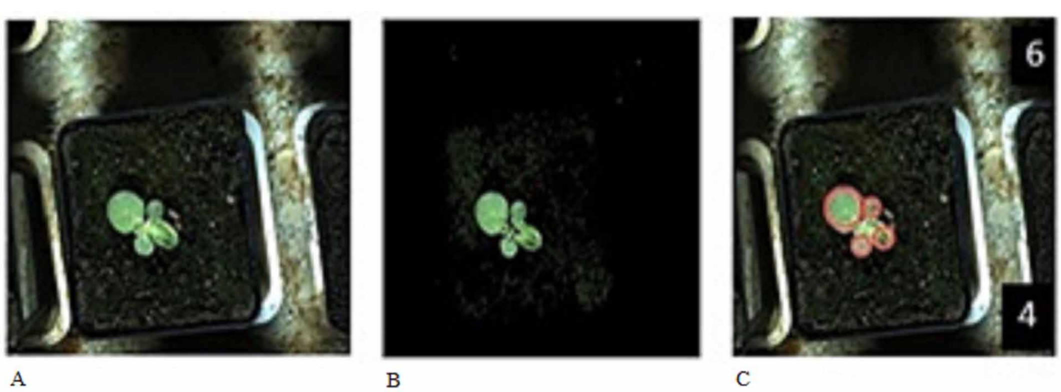
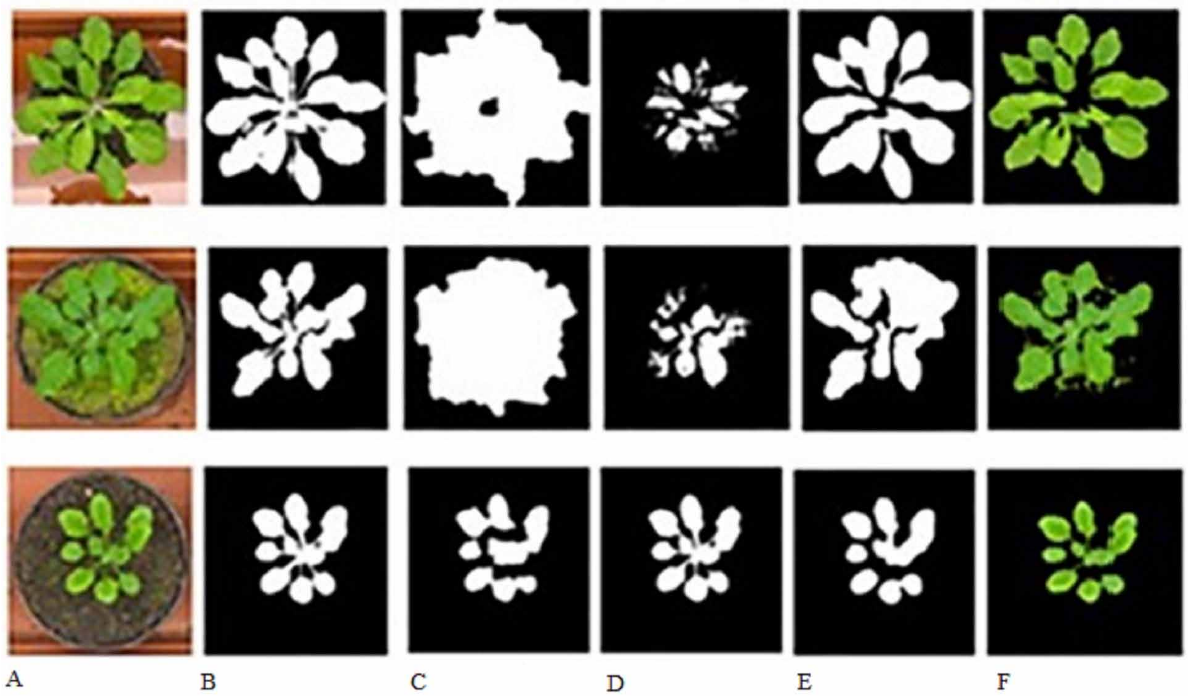


Figure 9. Plant region segmentation of individual plant images using various methods. (A) Original image, (B) Ground truth, (C) CoSand (2011), (D) SDSP (2013), (E) gPb-owt-ucm (2011), (F) Proposed method.



achieves overall segmentation accuracy of 94.5%, which is higher than the recent methods. Also, the standard deviation (4.5) of the proposed segmentation method is less when compared with other methods and thus maintaining the robustness of the proposed method. Table 2 shows the comparison of leaf counting performance of the two proposed leaf counting techniques. It is observed from Table 2 that the deep learning model (Proposed Fine-tuned VGG16) outperforms the CHT method in terms of leaf counting accuracy. Hence, it is used to compare the recent leaf counting methods and it is shown in Table 3.

It is observed from Table 3 that the fine-tuned VGG16 performs better than the recent methods. The |DiC| and DiC values of the proposed Fine-tuned VGG16 (deep learning) model are nearer to zero when compared to other methods. This shows that the leaf count per plant is found accurately when compared with other recent methods. But the recent methods have comparatively higher values of |DiC| and DiC for the CVPPP datasets. This shows that the performance of other methods is not consistent among various datasets. The standard deviation for DiC and |DiC| are less which indicates the proposed deep learning model is less biased towards overcounting or undercounting. This shows that the proposed deep learning model performs well by maintaining the robustness among different datasets.

Overall, the accuracy of proposed method in segmenting the plant region and counting the leaves is equally good for all the images in the dataset. The main advantages of the proposed method are: (i) independent of device and preprocessing software and (ii) applicable in various phenotyping platforms and environments with minimal modification.

Table 1. Performance comparison of plant region segmentation results for CVPPP datasets

Metric	Method	A1	A2	A3	ALL
FBD%	Nottingham	94.7 (1.5)	88.7 (16.2)	87.0 (25.0)	91.6 (16.0)
	Wageningen	94.9 (2.2)	95.2 (2.3)	91.6 (16.7)	94.0 (7.8)
	Hist_Quantize (2017)	93.2 (4.6)	94.3 (3.8)	81.3 (18.2)	91.7 (12.2)
	Graph-based (2018)	95.8 (3.1)	96.3 (2.1)	90.4 (10.9)	94.3 (6.8)
	Proposed Wavelet-based	96.1 (1.2)	92.5 (5.5)	92.6 (6.6)	94.5 (4.5)

Table 2. Performance comparison of two proposed leaf counting methods for CVPPP datasets

Method	Dataset	|DiC|	DiC
Proposed CHT-based	A1 A2 A3 ALL	2.8 (1.9) 1.8 (1.2) 1.7 (1.1) 2.3 (1.7)	-0.1 (3.4) -0.5 (2.1) -0.7 (1.9) -0.3 (2.9)
Proposed Fine-tuned VGG16	A1 A2 A3 ALL	0.81 (0.92) 1.88 (2.32) 1.25 (1.61) 1.16 (1.51)	-0.33 (1.19) -0.33 (3.04) 0.03 (2.04) -0.12 (1.91)

Table 3. Performance comparison of proposed leaf counting method with recent methods

Method	Dataset	|DiC|	DiC
Nottingham	A1 A2 A3 ALL	3.8 (1.9) 1.9 (1.7) 2.5 (2.4) 2.9 (2.3)	-3.5 (2.4) -1.9 (1.7) -1.9 (2.9) -2.4 (2.8)
Wageningen	A1 A2 A3 ALL	2.2 (1.6) 0.4 (0.5) 3.0 (4.9) 2.5 (3.9)	1.3 (2.4) -0.2 (0.7) 1.8 (5.5) 1.5 (4.4)
GLC (2016)	A1 A2 A3 ALL	1.27 (1.15) 2.44 (2.88) 1.36 (1.37) 1.43 (1.51)	-0.79 (1.54) -2.44 (2.88) -0.04 (1.93) -0.51 (2.02)
DCDN (2017)	A1 A2 A3 ALL	1.00 (1.00) 1.56 (0.88) 3.46 (4.04) 2.46 (2.73)	-0.33 (1.38) -0.22 (1.86) 2.71 (4.58) 1.42 (3.3)
Proposed Fine-tuned VGG16	A1 A2 A3 ALL	0.81 (0.92) 1.88 (2.32) 1.25 (1.61) 1.16 (1.51)	-0.33 (1.19) -0.33 (3.04) 0.03 (2.04) -0.12 (1.91)

FUTURE RESEARCH DIRECTIONS

There are few limitations in the proposed work. The segmentation accuracy is affected by the presence of shadow region in the plant images. This can be overcome by applying image enhancement before segmenting the plant region. Also, the leaf counting accuracy can be improved by designing new deep learning architectures.

CONCLUSION

A new method to segment the plant region and to count the leaves in the plant image is proposed in this chapter. The proposed method relies on the Wavelet transform and Circular Hough Transform. The proposed method is visually compared with existing tools and methods. The proposed method achieves better segmentation accuracy and counting accuracy than the existing methods or tools. The proposed method can be generalized since it can segment the plant region not only in Rosette Arabidopsis plants, but also for all the plants that contain green leaves. The leaf counting method can be applied to all the plants that contain elliptical or round shaped leaves. In order to enlarge the applicability of this proposed method in the field of plant phenotyping in future, the proposed method can be converted as an open source.

REFERENCES

An, N., Palmer, C. M., Baker, R. L., Markelz, R. C., Ta, J., Covington, M. F., & Weinig, C. (2016). Plant high-throughput phenotyping using photogrammetry and imaging techniques to measure leaf length and rosette area. *Computers and Electronics in Agriculture*, *127*, 376–394. doi:10.1016/j.compag.2016.04.002

Arbelaez, P., Maire, M., Fowlkes, C., & Malik, J. (2011). Contour detection and hierarchical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*(5), 898–916. doi:10.1109/TPAMI.2010.161 PMID:20733228

Cerutti, G., Tougne, L., Vacavant, A., & Coquin, D. (2011, September).A parametric active polygon for leaf segmentation and shape estimation. In *Proceedings of International Symposium on Visual Computing* (pp. 202-213). Springer. 10.1007/978-3-642-24028-7_19

Chéné, Y., Rousseau, D., Lucidarme, P., Bertheloot, J., Caffier, V., Morel, P., & Chapeau-Blondeau, F. (2012). On the use of depth camera for 3D phenotyping of entire plants. *Computers and Electronics in Agriculture*, *82*, 122–127. doi:10.1016/j.compag.2011.12.007

Choudhary, R., & Parmar, G. (2016, November).A robust image watermarking technique using 2-level discrete wavelet transform (DWT). In *Proceedings of 2016 2nd International Conference on Communication Control and Intelligent Systems (CCIS)*, (pp. 120-124). IEEE. 10.1109/CCIntelS.2016.7878213

De Vylder, J., Ochoa, D., Philips, W., Chaerle, L., & Van Der Straeten, D. (2011, October).Leaf segmentation and tracking using probabilistic parametric active contours. In *Proceedings of International Conference on Computer Vision/Computer Graphics Collaboration Techniques and Applications* (pp. 75-85). Springer. 10.1007/978-3-642-24136-9_7

De Vylder, J., Vandenbussche, F. J., Hu, Y., Philips, W., & Van Der Straeten, D. (2012). Rosette tracker: an open source image analysis tool for automatic quantification of genotype effects. *Plant Physiology*, 112.

Dellen, B., Scharr, H., & Torras, C. (2015). Growth signatures of rosette plants from time-lapse video. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *12*(6), 1470–1478. doi:10.1109/TCBB.2015.2404810 PMID:26684463

Giuffrida, M. V., Minervini, M., & Tsaftaris, S. A. (2016). *Learning to count leaves in rosette plants*. Academic Press.

Grand-Brochier, M., Vacavant, A., Cerutti, G., Kurtz, C., Weber, J., & Tougne, L. (2015). Tree leaves extraction in natural images: Comparative study of preprocessing tools and segmentation methods. *IEEE Transactions on Image Processing*, *24*(5), 1549–1560. doi:10.1109/TIP.2015.2400214 PMID:25667351

Huang, Y. J., & Lee, F. F. (2010). An automatic machine vision-guided grasping system for Phalaenopsis tissue culture plantlets. *Computers and Electronics in Agriculture*, *70*(1), 42–51. doi:10.1016/j.compag.2009.08.011

Illingworth, J., & Kittler, J. (1987). The adaptive Hough transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-9*(5), 690–698. doi:10.1109/TPAMI.1987.4767964 PMID:21869428

Kim, G., Xing, E. P., Fei-Fei, L., & Kanade, T. (2011, November).Distributed cosegmentation via submodular optimization on anisotropic diffusion. In *2011 International Conference on Computer Vision* (pp. 169-176). IEEE.

Ning, J., Zhang, L., Zhang, D., & Wu, C. (2010). Interactive image segmentation by maximal similarity based region merging. *Pattern Recognition*, *43*(2), 445–456. doi:10.1016/j.patcog.2009.03.004

Orlando, F., Napoli, M., Dalla Marta, A., Natali, F., Mancini, M., Zanchi, C., & Orlandini, S. (2011). Growth and development responses of tobacco (Nicotianatabacum L.) to changes in physical and hydrological soil properties due to minimum tillage. *American Journal of Plant Sciences*, *2*(3), 334–344. doi:10.4236/ajps.2011.23038

Praveen Kumar, J., & Domnic, S. (2017, March). A new histogram based quantization method to segment and count the rosette plant leaves. In *2017 IEEE International Conference on Current Trends in Advanced Computing (ICCTAC)* (pp. 1-5). IEEE.

Praveen Kumar, J., & Domnic, S. (2018). Image Based Plant Phenotyping using Graph Based Method and Circular Hough Transform. *Journal of Information Science and Engineering*, *34*(3), 671–686.

Scharr, H., Minervini, M., Fischbach, A., & Tsaftaris, S. A. (2014, July). Annotated image datasets of rosette plants. In *Proceedings of European Conference on Computer Vision* (pp. 6-12). Zürich, Suisse: Academic Press.

Scharr, H., Minervini, M., French, A. P., Klukas, C., Kramer, D. M., Liu, X., & Yin, X. (2016). Leaf segmentation in plant phenotyping: A collation study. *Machine Vision and Applications*, *27*(4), 585–606. doi:10.100700138-015-0737-3

Shen, W. Z., Zhang, C. L., & Chen, Z. L. (2007, August). Research on automatic counting soybean leaf aphids system based on computer vision technology. In *Proceedings of 2007 International Conference on Machine Learning and Cybernetics* (Vol. 3, pp. 1635-1638). IEEE. 10.1109/ICMLC.2007.4370408

Simonyan, K., & Zisserman, A. (2014). *Very deep convolutional networks for large-scale image recognition*. arXiv preprint arXiv:1409.1556

Tang, X., Liu, M., Zhao, H., & Tao, W. (2009, October). Leaf extraction from complicated background. In *Proceedings of 2nd International Congress on Image and Signal Processing, 2009.CISP'09* (pp. 1-5). IEEE.

Ubbens, J. R., & Stavness, I. (2017). Deep plant phenomics: A deep learning platform for complex plant phenotyping tasks. *Frontiers in Plant Science*, *8*, 1190. doi:10.3389/fpls.2017.01190 PMID:28736569

Yin, X., Liu, X., Chen, J., & Kramer, D. M. (2014, October).Multi-leaf tracking from fluorescence plant videos.I n *Proceedings of 2014 IEEE International Conference on Image Processing (ICIP)*, (pp. 408-412). IEEE. 10.1109/ICIP.2014.7025081

Zhang, L., Gu, Z., & Li, H. (2013, September). SDSP: A novel saliency detection method by combining simple priors. In *Proceedings of 2013 20th IEEE International Conference on Image Processing (ICIP)*, (pp. 171-175). IEEE. 10.1109/ICIP.2013.6738036

Chapter 6
Automatic Data Acquisition and Spot Disease Identification System in Plants Pathology Domain:
Agricultural Intelligence System in Plant Pathology Domain

Rajesh T. M.
Dayananda Sagar University, India

Kavyashree Dalawai
Dayananda Sagar University, India

Pradeep N.
Bapuji Institute of Engineering and Technology, India

ABSTRACT

Plants play one of the main roles in our ecosystem. Manual identification for the leaves sometimes leads to greater difference due to look alike. People often get confused with lookalike leaves which mostly end in loss of life. Authentication of original leaf with look-alike leaf is very essential nowadays. Disease identification of plants are proved to be beneficial for agro-industries, research, and eco-system balancing. In the era of industrialization, vegetation is shrinking. Early detection of diseases from the dataset of leaf can be rewarding and help in making our environment healthier and green. Implementation involves proper data acquisition where pre-processing of images is done for error correction if present in the raw dataset. It is followed by feature extraction stage to get the best results in further classification stage. K-mean, PCA, and ICA algorithms are used for identification and clustering of diseases in plants. The implementation proves that the proposed method shows promising result on the basis of histogram of gradient (HoG) features.

DOI: 10.4018/978-1-5225-9632-5.ch006

INTRODUCTION

Human race depends on plants directly or indirectly for its survival. What not we get from plants? Plants give us food, clothes, medicine, furniture and much more things. Healthy plants mean better quality of life for human being. Diseases in plants can decrease the production, increase the cost and might range to overall economic adversity of a produce if not alleviated suitably at initial phases et. al (Rao & Patel, 2012; Kanungo et al., 2016). The crops need planned nursing to distinguish the early indications in demand to avoid the feast of any plant infection, with low cost and better yield in manufacture. Employing trained agriculturists might not be reasonable particularly in isolated topographical regions. A steady plant monitoring is necessary to control the spread of a disease but its cost may be high and as a result, the producers often skip critical preventive procedures to keep the production cost low. Although, official disease recognition is a responsibility of professional agriculturists, low cost observation and computational assisted diagnosis can effectively help in the recognition of a plant disease in its early stages. A steady plant monitoring is necessary to control the spread of a disease but its cost may be high and as a result, the producers often skip critical preventive procedures to keep the production cost low. Although, official disease recognition is a responsibility of professional agriculturists, low cost observation and computational assisted diagnosis can effectively help in the recognition of a plant disease in its early stages. Gradually with technical and scientific advancement, more reliable methods through lowest turnaround time are developed and proposed for early detection of plant disease. Such techniques are widely used and proved beneficial to farmers as detection of plant disease is possible with minimal time span and corrective actions are carried out at appropriate time. In this chapter, we studied and evaluated existing techniques for detection of plant diseases to get clear outlook about the techniques and methodologies followed. Gradually with technical and scientific advancement, more reliable methods through lowest turnaround time are developed and proposed for early detection of plant disease. Such techniques are widely used and proved beneficial to farmers as detection of plant disease is possible with minimal time span and corrective actions are carried out at appropriate time. Since recent decades, digital image processing, image analysis and machine vision have been sharply developed, and they have become a very important part of artificial intelligence and the interface between human and machine grounded theory and applied technology. These technologies have been applied widely in industry and medicine, but rarely in realm related to agriculture or natural habitats. Many images used to develop new methods are collected under very strict conditions of lighting, angle of capture, distance between object and capture device, among others. This is a common practice and is perfectly acceptable in the Plant Disease Detection using Computer Vision early stages of research. Manual identification for the leaves sometimes leads to greater difference due to look alike. People often get confused with lookalike leaves which mostly end in loss of life. Authentication of original leaf with look-alike leaf is very essential nowadays. However, in most real world applications, those conditions are almost impossible to be enforced, especially if the analysis is expected to be carried out in a non-destructive way. Computer vision can propose a substitute answer in plant nursing and such an approach might help in predicting the diseases at early stages. Knowledge learning on images has been proved to be pioneer in early identification of diseases. Data clustering using k-mean method is a predominant field of investigation in pattern recognition. K-mean remains termed as the most robust and popular method in hard clustering. The enactment of the recommended technique has been paralleled with the existing models by means of synthetic datasets. This chapter presents an approach for plant leaf image segmentation by applying linear k means algorithm. The segmentation process presents a clustering mechanism for high resolution images in order to improve the precision and processing time. Plant image, however, always contain complicated background objects that interfere with the examination process and must be removed from the image prior to species classification. K means clustering is applied at the first level of segmentation to detect the structure of the plant leaf. The features are then extracted and classification is done by Multiclass SVM technique. The database is created with different kinds of leaves and other medicinal purpose leaves. For the execution of test leaf a test image is captured and the parameter is tested. With this the leaf which gives the closest match will be labelled. The following sections describe some of the terms associated with the chapter.

K-means is unique and the easiest unsubstantiated clustering and learning algorithms. The procedure surveys a modest and informal way to categorize assumed data set through a fixed quantity of clusters (assuming k clusters). The elementary idea is to express k centers, one for individual cluster. The K-means clustering remains a kind of unsubstantiated learning, which is used when you have uncategorized data (i.e., data without defined boundaries or groups). The unsubstantiated learning is a type of machine learning algorithm used to pull implications from datasets comprising of input data without categorized responses. The most common unsubstantiated learning system is the study of clusters, which is used for experimental data analysis to find unseen patterns or categories in data.

The fundamental optimization methods are helpful in finding the optimum solution or unconstrained maximum or minimum of differentiable and continuous functions. Optimizations a technique of making something like a decision, system, or design as fully effective, perfect or functional as far as possible specially the mathematical derivations of finding the maximum or minimum of a function. Optimization difficulties can be of two types conditional on whether the variables are discrete or continuous. Accordingly, there are two diverse types of optimization approaches used widely today (Wu et al., 2016; Sankaran et al., 2010). Firstly, is deterministic method and secondly is stochastic methods. Deterministic methods involve specific procedure or rules for finding solution to a problem. These methods have remained effectively applied for many manufacturing enterprise problems. In stochastic methods involves accordance with the translation rules related to probability used to model un-deterministic nature of the real world. These methods are gaining value owing to the convinced possessions which deterministic algorithms do not.

The main goal line of optimization is to gain the finest conceivable solution with respect to a group of concerned and prioritized constraints or criteria. This decision-making procedure is known as optimization. Optimization includes maximizing the factors such as reliability, efficiency, productivity, utilization strength and longevity. For instance, code optimization is the procedure of modifying code to refine quality and efficiency of code. A code can be optimized to make it smaller in size, taking small space thus executing faster with lesser input and output tasks. Therefore, identification and verification method is chosen to identify diseases in the leaf using digital images. The automated identification and verification system is a significant procedure where a separate try to study somewhat from other persons to recover information about certain occurrence.

Motivation

For increasing growth and productivity of crop field, farmers need automatic monitoring of disease of plants instead of manual. Manual monitoring of disease does not give satisfactory result as naked eye observation is old method requires more time for disease recognition also need expert hence it is non-effective. In some countries, consulting experts to find out plant disease is expensive and time consuming due to availability of expert. Irregular checkup of plant results in growing of various diseases on plant which requires more chemicals to cure it also these chemicals are toxic to other animals, insects and birds which are helpful for agriculture. Automatic detection of plant diseases is essential to detect the symptoms of diseases in early stages when they appear on the growing leaf and fruit of plant. So, in this chapter, a modern technique to find out disease related to leaf is introduced. To overcome disadvantages of traditional eye observing technique, digital image processing technique for fast and accurate disease detection of plant is an alternate solution.

ANALYSIS ON DIFFERENT TYPES OF PLANT DISEASES AND THE WORK DONE ON IN THE PLANT PATHOLOGY DOMAIN

Bacteria and fungi are the most common parasites causing plant disease. Fungi usually produce the spores which, when carried to a plant, can begin an infection. Bacteria can attack living plants and can cause the plant disease. These diseases can be carried from plant to plant by wind, water, insects etc. These diseases occur primarily on leaves, but some can even spread to stem and fruits. Leaf disease is the most common diseases of most plants. Although the leaf diseases are described under several symptom types, yet the difference is not clear and there are many names for leaf diseases. Below is the list of diseases that are caused by both bacteria and fungi.

1. **Bacterial Spots:** The most common symptom of bacterial disease is leaf spot. These appear on leaves, fruits and stem. If the spot appears and advances rapidly the disease is considered as blight.
2. **Cankers**: Primarily Pseudomonas and Xanthomonas cause canker disease of a stone fruit tree and canker disease of citrus respectively. Canker symptoms appear on trunks, stem, twigs and branches.
3. **Bacterial Galls**: Galls can be produced by the genus Agrobacterium and certain species of Arthrobacter, Pseudomonas, Rhizobacter and Rhodococcus. Agrobacterium tumefaciens, A. rubi and A. vitis alone are responsible for galls in over 390 plant genera worldwide. Gall tissue is composed of disorganized, randomly proliferating cells that multiply in the intercellular (between the cells) spaces in the vicinity of the wound.
4. **Bacterial Vascular Wilts**: These primarily affect the herbaceous plants such as vegetables, field crops, ornamentals and some topical plants. The symptoms include wilting and death of above ground parts of the plants.
5. **Bacterial Soft Rots**: Primarily the bacteria that cause soft rots in living plant tissue include Erwinia spp., Pseudomonas spp., Bacillus spp. and Clostridium spp. Soft rots attack a large number of hosts and are best known for causing disease in fleshy plant structures both above and below the ground.
6. **Bacterial Scabs**: These primarily infect belowground parts of plants such as potatoes
7. Rusts: These often produce spots similar to leaf spots, but are called “pustules”. Rust pustules are usually raised above the leaf surface. The leaf withers and dies rapidly.
8. **Powdery Mildew**: This is superficial, white to light grayish, powdery to mealy growth on leaves, but may occur on stem and fruits.
9. **Downy Mildew**: The symptoms are pale yellow green to yellow areas in the upper leaf surface. Light gray to purple moldy growth on under surface of the leaf.

Rao and Patel (2012) have explained the use of computers and computer technologies that are explored in various fields as science, engineering, medicine, commerce, agriculture. Significant progress in the area of computer vision and image processing (CVIP) has led a real-world applications in various fields. They have considered one of the applications to identify the fungal diseases in the plant from visual symptoms. Statistical features using block-wise, Gray Level Co-occurrence Matrix(GLCM) and Gray Level Run Length Matrix(GLRLM) are extracted from image samples. They have used Nearest Neighbor classifier using Euclidean distance to classify images into partially affected, moderately affected, severe affected and normal.

Kanungo et al. (2016) have worked on automatic detection and classification of plant leaf disease. They have used basic image processing techniques for enhancing the image. They have also used image segmentation method to separate the gray scale image into different parts based on the region of interest. With this they have used Genetic Algorithm for color image segmentation. Genetic algorithm belongs to the evolutionary algorithms which generate solutions for optimization problems. They have used SVM (Support Vector Machine) with their proposed algorithm. They have achieved overall 97% of accuracy.

The authors, Wu et al. (2016), have given a brief analysis study on various image processing techniques that can be used for automatic plant disease identification and detection. They have considered various diseases caused in plants like late blight, and Canker.

The scholars, Sankaran et al. (2010), have given brief survey on plant leaf disease classification using image processing. They have given a brief idea about the classification techniques like K-Nearest Neighbor(KNN), Radial Basis Function(RBF), Probabilistic Neural Networks(PNN), Support Vector Machine(SVM). They have explained each of these techniques with its advantages and disadvantages. They have concluded that the not all techniques are equally efficient, but few are best algorithms like SVM (Support Vector Machine) for identifying only bacterial spots.

Authors, Bonanomi et al. (2010), have given importance to detect and classify the plant leaf disease using image processing techniques. They have used Artificial Neural Networks methods for classifying the images. They have also used K-Means clustering algorithm to cluster the images into diseased and normal plant images.

Scholars, Al-Hiary et al. (2011), have worked on identifying the plant leaf disease based on the lesion features which is implemented in the mobile phones. They basically named it as Spot Recognition System (SRS). They converted the color image in to binary image and then the spot disease is recognized using the mobile camera which will be connected to the GPRS or internet. Later the results can be further transferred through different communication channel like Bluetooth, Wifi etc. they have achieved 90% of accuracy in detecting the plant leaf spot disease.

The main aim of the work (Sladojevic et al., 2016) is to identify and classify the disease accurately from the leaf images. They have considered Powdery Mildew and Downey Mildew disease which can cause great loss for grape fruit. The major and minor axis of the leaf is extracted and then forwarded to Artificial Neural Network classifier to classify the affected and unaffected leaf.

The authors, Korkut, Göktürk, and Yildiz (2018) have analyzed the various techniques that can be used to detect and classify the diseased plant leaves. They have worked on multiple plant leaf databases like soyabean leaves, cotton leaves, wheat leaves and grape leaves. The classifier algorithm/classifier used on soaybean leaf are SIFT (Scale-Invariant Feature Transform) and SVM (Support Vector Machine) classifier and they have achieved 93% of accuracy in classifying the diseased leaf. For cotton leaves they have used PCA (Principle Component Analysis) and KNN (K- Nearest Neighbour) for which they have got 95% of accuracy. Similarly, for wheat and grape leaves they have used PCA (Principle Component

Table 1. Analysis on different techniques used by various researchers on different types of plant diseases

Sl. No	Plant Disease	Algorithm/ Methods Used	Results
1	Fungal disease based on visual symptoms (Rapo & Patel, 2012)	Gray Level Co-occurrence Matrix (GLCM) and Gray Level Run Length Matrix (GLRLM)	91.37% and 86% using GLCM and GLRLM features respectively.
2	Bacterial Disease, Sun burn disease, Early Scorch and Fungal disease (Kanungo et al., 2016)	Genetic algorithm and Support Vector Machine	97% of accuracy
3	Bacterial spot disease (Al-Hiary et al., 2011)	Lesion Features and Spot Recognition System (SRS)	90% of accuracy
4	Powdery Mildew and Downey Mildew disease (Sladojevic et al., 2016)	Artificial Neural Network (ANN)	Expected 94% of accuracy
5	Sheath Blight (Korkut, Göktürk & Yildiz, 2018)	SIFT (Scale-Invariant Feature Transform) and SVM (Support Vector Machine)	93% of accuracy
6	Cotton Wilt (Korkut, Göktürk & Yildiz, 2018)	PCA (Principle Component Analysis) and KNN (K- Nearest Neighbor)	95% of accuracy
7	Brown rust (Korkut, Göktürk & Yildiz, 2018)	PCA (Principle Component Analysis), BPNN (Back Propagation Neural Network)	96% of accuracy
8	Downy Mildew (Korkut, Göktürk & Yildiz, 2018)	PCA (Principle Component Analysis) and K-means	76% of accuracy

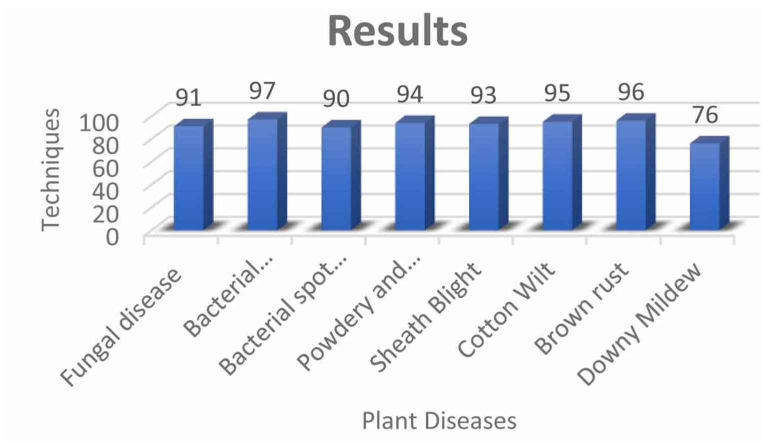
Analysis), BPNN (Back Propagation Neural Network) and K-means methods respectively, and they have achieved 96% and 76% accuracy. The below Table 1 explains about the analysis results of various techniques versus various diseases.

A NOVEL APPROACH FOR SPOT DISEASE IDENTIFICATION AND VERIFICATION SYSTEM IN THE PLANT PATHOLOGY DOMAIN

The proposed approach consists of four crucial phases. The pre-processing of dataset is considered as the first stage of data clustering, which is followed by feature extraction. The second stage involves identification and verification of images. The following are the steps to detect diseases in plants.

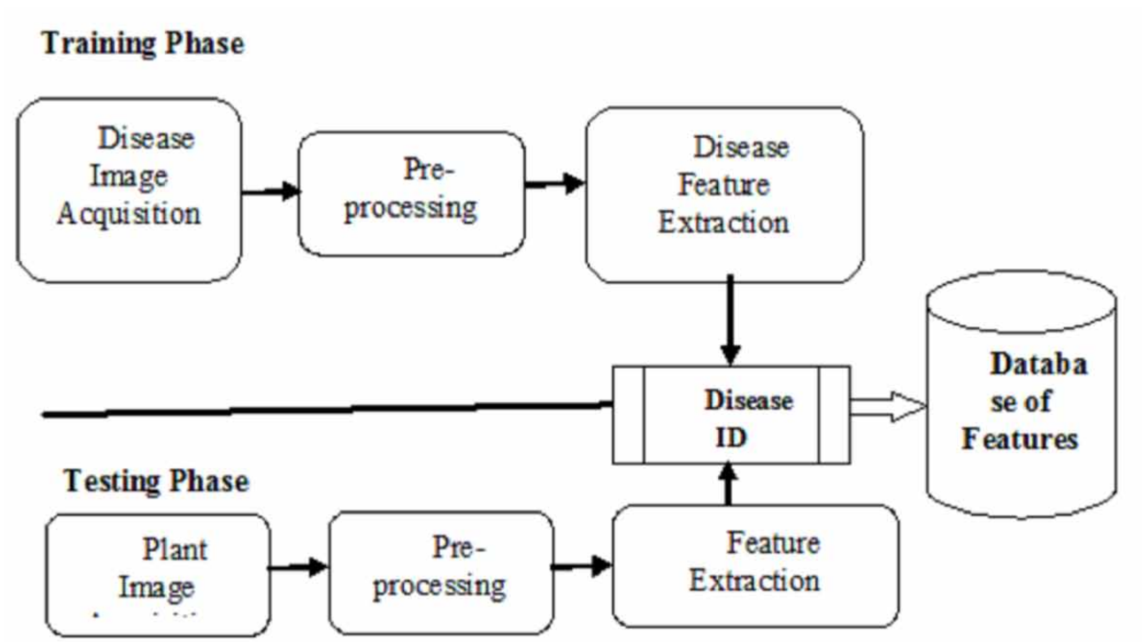
1. **Preprocessing**: In this step, noise is removed from the image obtained and images are resized as per requirement.
2. **Feature Extraction**: In this step, the features of the image are extracted by following stepwise procedures:
    a. Spot feature extraction, where clustering based on colors (green and others) is carried out.
    b. Clustering based on foreground (spots) and background (green color) features.
    c. Using kernel method, spots are computed depending on weight and numbers of the spots identified.
    d. Principal Component Analysis and Independent Component Analysis are used for translation of points in feature extraction.
3. **Identification**: Matching the features of images is done in this step.

Figure 1. Graphical analysis of results and accuracy of different techniques versus different plant diseases



4. **Verification**: In this step, verification is done using the Error Estimation Ratio (EER), False Acceptance Ratio (FAR) and False Reject Ratio (FRR). In this step, the HoG features are used to get accurate results. Figure 1 explains the architecture of the Identification and Verification of Plant Diseases (Tichkule & Gawali, 2016). It consists of two phases as Training and Testing phases as shown in Figure 2. Flowchart for the proposed approach for depicting various techniques used in this chapter for implementation is shown in Figure 3.

Figure 2. Block diagram of the proposed approach for identification and verification of plant diseases (Adapted from (Rajesh et. al., 2019))



The steps involved are

1. **Training**: In training all the collected images are trained to the model and all six features are extracted and stored in the database. The system design mainly consists of
    a. Image Acquisition
    b. Image Preprocessing
    c. Image Segmentation
    d. Feature Extraction
    e. Training
    f. Classification using Multiclass SVM
2. **Classification**: After training, the SVM will classify the given new input as which type of disease is affected. The system design mainly consists of
    a. Image Acquisition
    b. Image Preprocessing
    c. Image Segmentation
    d. Feature Extraction
    e. Training
    f. Classification using Multiclass SVM

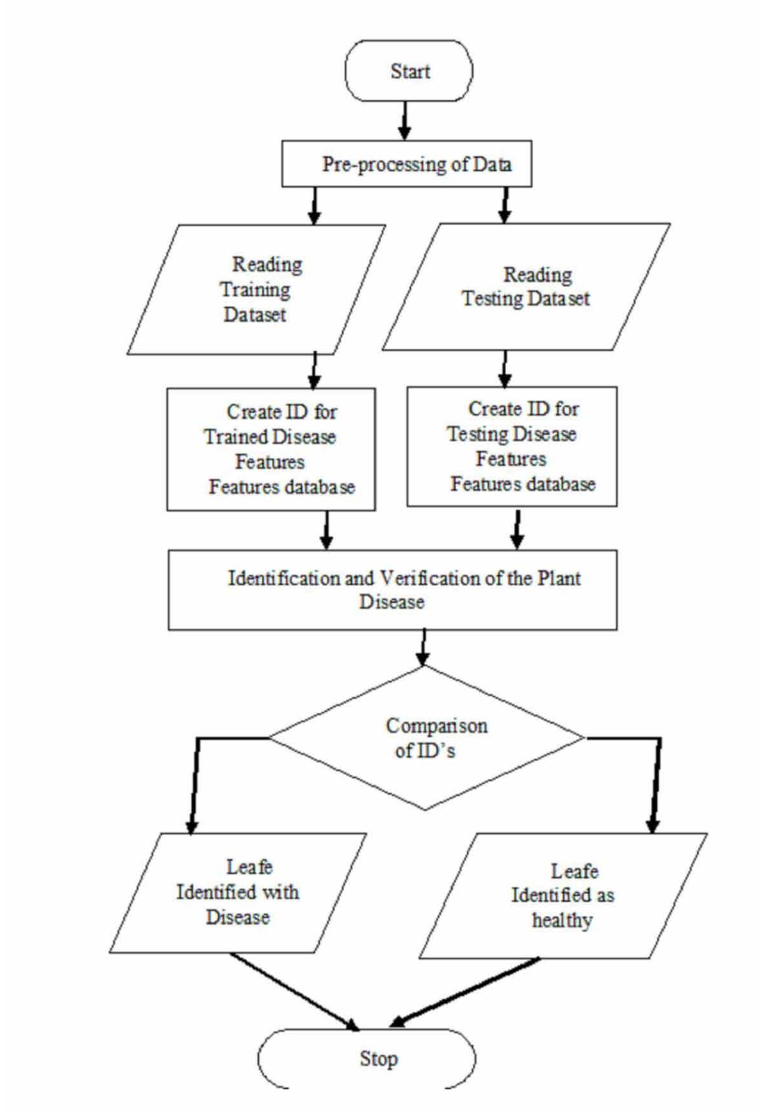
Preprocessing of Images

The preprocessing stage involves spot estimation, which is a mechanism to detect the disease spot in the leaf from the acquired leaf image. Once the image is obtained, spot value can be estimated. Prior to detecting spot value of the leaf, preprocessing of the image needs to be done. The preprocessing steps involve monochromatic image, noise reduction in the image, and width normalization. After the above steps, the spot estimation gives better results. The following are the preprocessing steps

Grey Level Image

One of the utmost common problems in digital image processing applications is segmentation (Bonanomi et al., 2010; Al-Hiary et al., 2011; Sladojevic, Gokturk & Yildiz, 2018; Tichkule & Gawali, 2016). The segmentation involves the separation of components or layers in the image i.e. the ability to identify and to segregate the objects from the background. On the basis of application, the application of segmentation is proved to be crucial to obtain reliable results. Researchers working on image segmentation mainly divide segmentation into three basic steps, namely: image segmentation, region labelling and selection. In our approach the colored image is converted to monochromatic images and then noise is removed from the grey image. The colored images are converted to monochromatic (black and white) images using Niblack algorithm. The grey level images can be obtained using Paint tool or in MATLAB from RGB to gray as shown in Figure 4.

Figure 3. Flowchart depicting the plant disease identification implementation
(Adapted from (Rajesh et. al., 2017))



Noise Reduction in the Images

The noise is termed as the unnecessary data in the dataset which dilutes the result to a large extent. Many researchers even use rough set or many advanced toll to reduce noise in the dataset. Here the dataset consists of images of leaves. After images are obtained, sometimes the noise gives incorrect clusters. So here median filter is used to reduce noise in the images. The median filter is basically a nonlinear digital filtering mechanism which removes irrelevant image portions from the main image. Such techniques are typically pre-processing techniques to improve the processing.

Figure 4. Original RGB image
(Adapted from (Rajesh et. al., 2019))

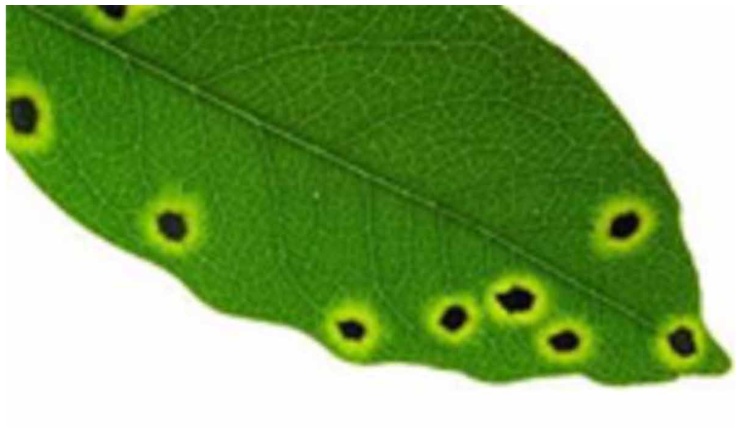


Figure 5. Grey image
(Adapted from (Rajesh et. al., 2019))

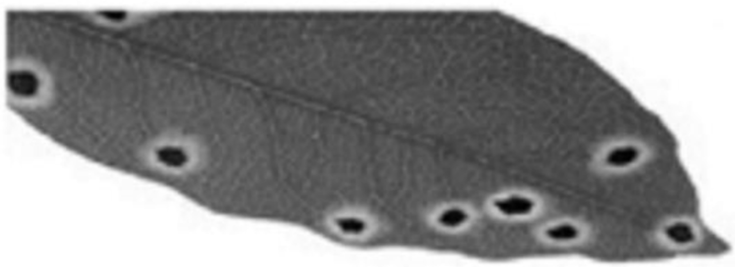


Figure 6. Noise image
(Adapted from (Rajesh et. al., 2019))

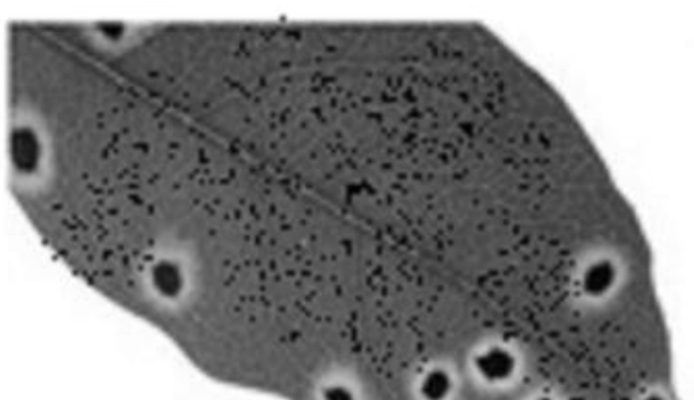
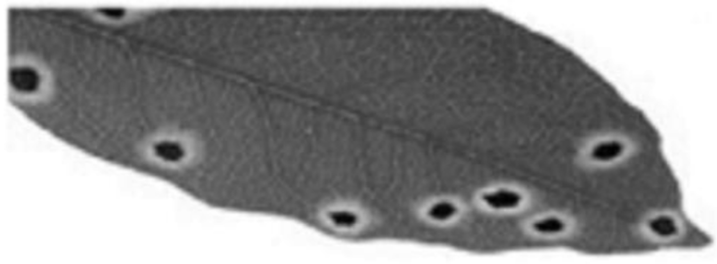


Figure 7. Noise free image
(Adapted from (Rajesh et. al., 2019))



Normalization of Width

The normalization is the procedure which changes the range of pixel/grid size of the image intensity values. Sometimes images get insufficient or too much contrast due to glare of light. These components need to be balanced by normalization techniques. It is otherwise called difference widening or histogram widening or vigorous range extension. The main objective of dynamic range development is to get the image or pointer into a given assortment which is acquainted to the intelligences, therefore it is termed as normalization. In order to maintain the uniformity in aspect size ratio for all the images which will be presented for testing as an input in disease identification and verification system, the monochromatic noise removed image is resized with the same size ratio. In our work we have maintained 250 * 250 sizes for all the images and we have used automating resizing MATLAB technique to gain this result. The Figure 8 shows the original image whereas 9 shows the normalized image.

Figure 8. Grey image
(Adapted from (Rajesh et. al., 2019))

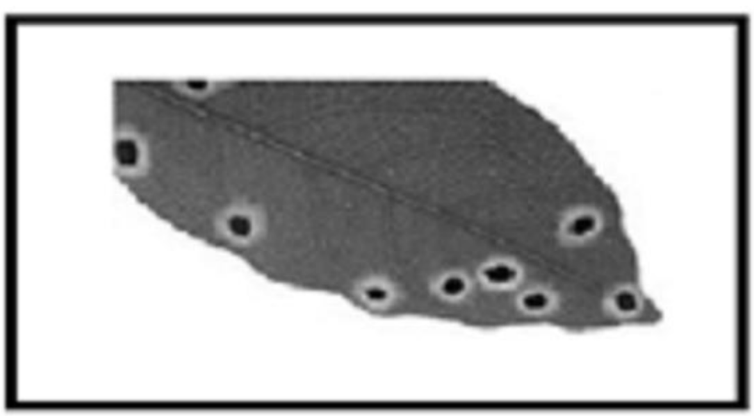
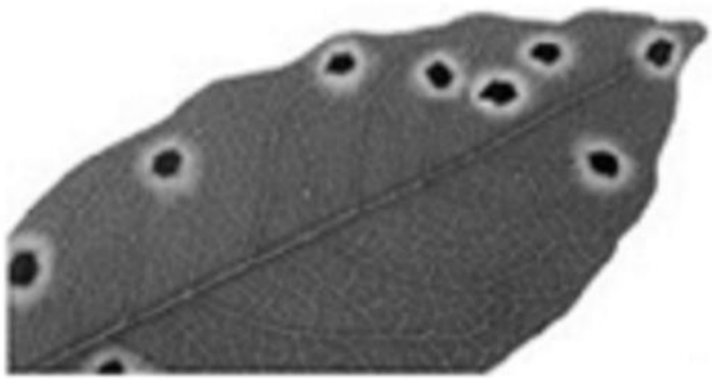


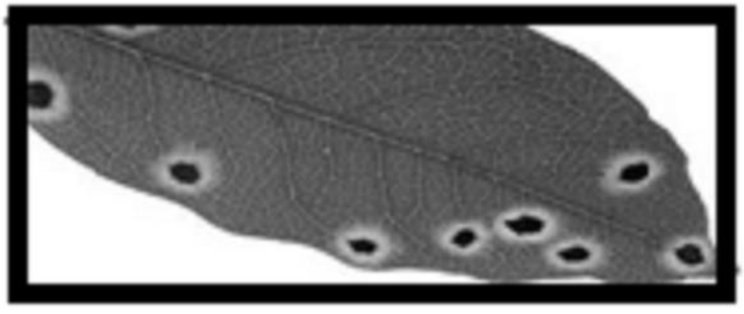
Figure 9. Resized image
(Adapted from (Rajesh et. al., 2019))



Cropping

Cropping is the crucial procedure of image processing in which the unwanted images are cropped out to relevant image grid. This procedure is important to elevate the performance of disease detection. In our work we have used the automatic cropping method which will crop the leaf by bound the rectangle over the leaf automatically and this might result in loss of data. So using region of interest (ROI) method along with the automatic cropping yields us in getting the good and accurate cropped leaf images. The Figure 8 shows the original image whereas Figure 10 shows the cropped image. In many of the pattern recognition problem, it is assumed that the data will be unimodal Gaussian structure, however this is not the case every time. The unimodal refers to the local maxima in a chart (Tichkule & Gawali, 2016; Ashourloo et al., 2016; Ashour et al., 2015; Rao & Patel, 2013; Kanungo et al., 2016; Krishna & Sao, 2016). It is found that data is mostly multimodal. In order to have clear understanding about the data distribution of the training set it is recommended to use multiple Gaussian components instead of single Gaussian envelope. This can be achieved through the mixture of many Gaussian distributions which is superposition of several Gaussian expressed in Equation 1 as follows.

Figure 10. Cropped image
(Adapted from (Rajesh et. al., 2019))



Spot Detection

$$A(y)=\sum_{m=1}^{m}\pi_m\eta(x/\mu_m,\Sigma_m) \tag{1}$$

The parameter 'm' in above Equation 1 is termed as mixing coefficient and above equation is termed as a combination of multivariate Gaussians, where $\eta(y/\mu_m, \Sigma_m)$ and each $\eta(y/\mu_m, \Sigma_m)$ are called as the components of the combination. The mixture is expressed in Equation 2 as given by

$$\eta(x/\mu_k,\Sigma_k)=\frac{1}{2\pi^{D/2}\,|\,\Sigma\,|^{1/2}}e^{-\frac{1}{2}(y-\mu_m)^T\sum_m^{-1}(y-\mu_m)} \tag{2}$$

Equations 1 and 2 have their own expected values mean (μ_m) and covariance matrix (Σ_m).

If mutual edges of Equation (1) are integrated with respect to 'y' and both a(y) and the individual Gaussian mechanisms are standardized, then $\sum_{m=1}^{m}\pi_m=1$ is obtained. Also, from the probability law it is required that a(y) ≥ 0, together with $\eta(y/\mu_m, \Sigma_m) \geq 0$, which implies m ≥ 0 for all k. Combining this with Equation (1) 0≤m≤1 is obtained. From the sum and product rules, the bordering compactness is expressed in equation 3 as given by

$$a(y)=\sum_{m-1}^{m}a(y)a(y/m) \tag{3}$$

Equation (1) in which $m=a(m)$, which is nothing but the past probability of selecting k^{th} component. Also, the mass η(y/μm, m) = a(m/y) is the probability of x trained on the parameter k. As per Bayesian theorem, the posterior probability a(m/y) is given in Equation 4, 5 and 6 as following.

$$y(Z_k)=p(k/x) \tag{4}$$

$$y(Z_k)=\frac{p(k)p(x/k)}{\Sigma lp(l)p(x/l)} \tag{5}$$

$$y(Z_k)=\frac{\pi_k\eta(x/\mu k,\sum k)}{\Sigma l\eta(\mu l/\sum l)} \tag{6}$$

Normally, the Gaussian combination distribution is described by the constraints π, μ and summation Σ, where the representation π ≡ π1, π2, . ….. ., πK, μ ≡ μ1, μ2, . …. ., μK and Σ ≡ Σ1, Σ2, . ….. ., ΣK is used. In order to estimate the parameters of μ, Σ and π we recommend to use Expectation-Maximization (EM) algorithm. In statistics, expectation-maximization algorithm is an algorithm where iteratively determined likelihood or determined a posteriori (MAP) is estimated. The exemplary depends a lot on unnoticed latent variables (Rajesh & Aradhya, 2015). The EM algorithm iteratively calculates an expectation (E) step, which generates the purpose for the anticipation of log of likelihood and a maximization (M) step, which computes parameter capitalize on the expected log-likelihood calculated on the E step. The obtained parameters are used to calculate the distribution of the latent variables in the iterative E step.As EM algorithm requires initial seed values of μ, Σ and π and they are chosen by running K-Means Clustering algorithm with 'k' as number of clusters where 'k' is the Gaussian mixture components. The expectation (E) step and the Maximization (M) step alternatively update the values of these parameters until a convergence criterion is reached. The EM algorithm is summarized as follows:

1. K-means clustering is carried on the entire data and 'k' number of clusters are obtained. From the k clusters parameters μ, k, Σk and are estimated as initial values and log likelihood function is evaluated using Equation 6.
2. In expectation step, using the current parameter values the responsibilities are evaluated using Equation 6).
3. 3. In maximization step, the parameters are re-estimated using the current responsibilities by using Equation 7, 8 and 9.

$$\mu k = \frac{1}{Nk}\sum_{n=1}^{N} y(Znk)U_m \tag{7}$$

$$\sum K = \frac{1}{Nk}\sum_{n=1}^{N} y(Znk)U_m(X_n - \mu k) \tag{8}$$

$$\pi k = \frac{Nk}{N} \tag{9}$$

4. 4. Once the new parameters are estimated they are evaluated using log likelihood function in Equation 10 as follows:

$$\mathrm{p}\left({}^{x}\!/\!_{\mu}, \Sigma, \pi \right) = \sum_{n=1}^{N}\sum_{k=1}^{k} \pi \,\mathrm{kn}(\mathrm{xn}/\mu\mathrm{k}, \Sigma\mathrm{k}) \tag{10}$$

In the current work, the achieved means of k clusters is used for estimating the spot of the image obtained. The mechanism of selecting assessment is extremely particular in nature. Therefore, we fix the cost of equal to number of connected components in the leaf image. The centroid of respective cluster is frozen as entrant points to estimate the Spot value of the leaf. After detecting the Spot value of the obtained image, in order to recognize the accurate spot in the image we have used nearest neighbor interpolation technique. In nearest neighbor interpolation, the spot value detected using EM algorithm is subjected to recognize the accurate spot in the image. The block uses the value of predicted value and nearby translated pixel values to correct the spot for the output pixel values. The output matrix which is created to recognize the accurate spot in the leaf image replaces each input value with the translated value nearest to it and the final output will be recognized with the accurate spot in the image. The next section describes the second step i.e. feature extraction.

Feature Extraction for Identification of Disease in Plants Using PCA

Principal Component Analysis (PCA) concept is widely castoff to decrease the dimensionality of data and also, it is used to extract the significant representative of a set of feature vectors. In PCA, all significant representatives are called as principal components and these principal components satisfied the condition of orthogonality. Kernel Principal Component Analysis (KPCA) is an extension of PCA; it extracts the features from non-linear subspace. In KPCA, the kernel said that solve the problem by increasing the dimensionality of the data. In our research work, these two concepts have been implemented to identify the spot disease in the plants. Following paragraphs describe PCA and KPCA in detail.

Principal Component Analysis (PCA)

PCA describes a calculated procedure that converts a quantity of perchance correlated variables into a lesser number of uncorrelated variables called as principal components. The first principal component presents large inconsistency in the data as far as conceivable, and each succeeding component of PCA determines the much of the remaining variability as far as possible. This confirms that, the first principal component has best representative of features. If first principal component could not represent the discriminative features, then succeeding principal components are selected until best representative of data.

Kernel Principal Component Analysis (KPCA)

KPCA is an improved version of PCA approach and it can also be assured as simplification and nonlinear description of principal component analysis. To achieve KPCA, first step is to transform the contribution data x from the unique input space R into a higher-dimensional feature space F with the nonlinear transmute $\Theta(x){:}\rightarrow F$. The map Θ and the space F are resolute implicitly by the excellent of a kernel purpose K, which calculates the dot creation amongst two input Instances x and y mapped into F via $K(x,y) = \Theta(x)\ \Theta(y)$. The kernel matrix is first integrated in Kernel PCA, which can be measured as the approximation of the covariance matrix of the new feature vector. This kernel PCA extracts the features from non-linear subspace, after kernel procedure, the data becomes linear and then PCA procedures are followed. In Kernel PCA, vectors are mapped through functions and kernels solve the problem by increasing the dimensionality of data (Aradhya, 2007). In this research work, the principal component analysis (PCA) procedure has been used for disease recognition task. PCA is a one of the statistical technique to highlight the similarities and differences exist in the given data. The projection vectors decide the dimension of the PCA features. From the set of training images, the PCA basis vectors are calculated, which helps to identify the principle components from the input data. Initially, the average image (Avg) is calculated by computing the mean of all training images (A1, A2, .., AN) as shown in Equation 11. The average image (Avg) is subtracted from the training image to get training data sample matrix (Tr) as shown in Equation 12.

$$AVG = \frac{1}{N}\sum_{n=1}^{N} Ai \tag{11}$$

$$T=A-AVG \tag{12}$$

The covariance matrix C is calculated from the training data sample matrix. Since the covariance matrix is square matrix, the Eigen values $\lambda_i(i=1\ldots N)$ and corresponding Eigen vectors $U_i(i=1\ldots N)$ were calculated.

The smallest Eigen value represents smallest information and largest Eigen value represents more information. In PCA, the Eigen vectors of largest Eigen value is selected, because Eigen vectors conforming to major Eigen value represents the greatest principal component but the smallest Eigen value gets less significance. The K number of projection vectors decides the k number of Eigen vectors to be considered which gives projection matrix (Uk). The feature matrix of training images (Ftr) is calculated by transposing the projection matrix and training data samples which is shown in Equation 13.

$$C = \frac{1}{N}\sum_{i=1}^{N} Tr_i Tr_i^T \tag{13}$$

KPCA extracts the features from nonlinear subspace. It preserves the subspace of the patterns and discards remaining space. Whereas PCA finds the patterns in linear subspace of lower dimensional data but KPCA finds the patterns in nonlinear subspace of high dimensional data. The first step in KPCA is to convert the unique input space into higher dimensional features space. The polynomial kernel is developed by using Equation 14.

$$k(i,l) = \sum_{i=1}^{N}\sum_{l=1}^{N} Tr(i,1:N)Tr(l,1:N) \tag{14}$$

After generating the kernel matrix, the centralization concept is applied and which can be considered as covariance matrix. The principal components are identified by performing PCA on centralized kernel matrix. The Eigen values $\lambda_i(i=1\ldots N\text{-}1)$ and consistent Eigen vectors $U_i(i=1\ldots N)$ are calculated from the centralized kernel matrix. The feature matrix of training samples is obtained by transposing of k number of Eigen vectors $U_k=U_i(i=1\ldots k)$ and input data. The next section describes the algorithms used in our proposed approach.

EXPERIMENTAL SETUP

This section is dedicated for the experimental set up for the proposed problem. It includes MATLAB R2017a on a PC having Intel I5 Processor with CPU T5800, processing capacity of 2 GHz, memory of 4 GB RAM and Microsoft Windows 2010 OS. The objective of the article is to progress the bunch centers as well as calculating the supreme suitability value by means of elitist-based PCA termed ICA. In instruction to assess the presentation of the projected system, many images available from sources like websites are considered. This dataset consists of 100 individual spot disease leaf images and each consists of 50 samples. Figure 11 shows the sample spot disease identified leaves.

In all experiment in this section, the population size and total evaluation number considered the same for fair simulation. Population size was 300. In order to conduct experimentation, we have created 30 dataset spot diseased leaves. We had taken 10 leafs from each 30 in dataset constituting 300 leaves. We have applied our proposed K-mean algorithm on all 300 leafs to detect the spot of the leaf. Results obtained on some sample images based on the proposed technique are as shown in Figure 13. The result obtained reveals that though K-mean algorithm is a competent classifier but ICA algorithm shows better result. For the analytic discussion, the K-mean algorithm is implemented and later integrated with PCA using the best or elite solutions form the population in order to develop the ICA-K mean algorithm. In the simulation, the value of k is kept at the range of 5 to 50 and d is set to 0.1. The rate of elitism considered was 0.1 and the value of teaching factor is 1.

In mandate to study the efficiency of the projected model, once the Spot is detected we calculate its value and once again we have applied our proposed model on spot detected on leaves to detect how much of spot are still present. The Figure 14 shows the 3 clusters obtained while identification of the spot on the leaves. In the cluster 1, only two clusters are considered in which the background is bifurcated

Figure 11. Dataset of leaf images for spot identification
(Adapted from (Rajesh et. al., 2019))

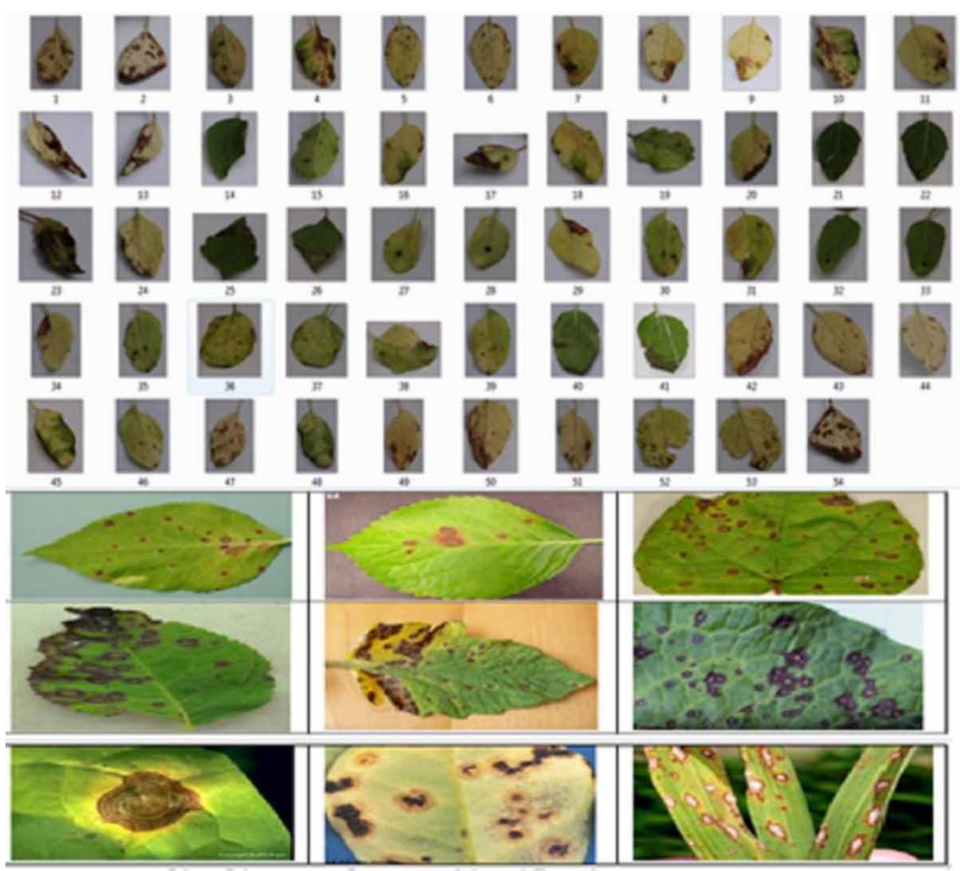


Figure 12. Dataset of healthy leaf images
(Adapted from (Rajesh et. al., 2019))

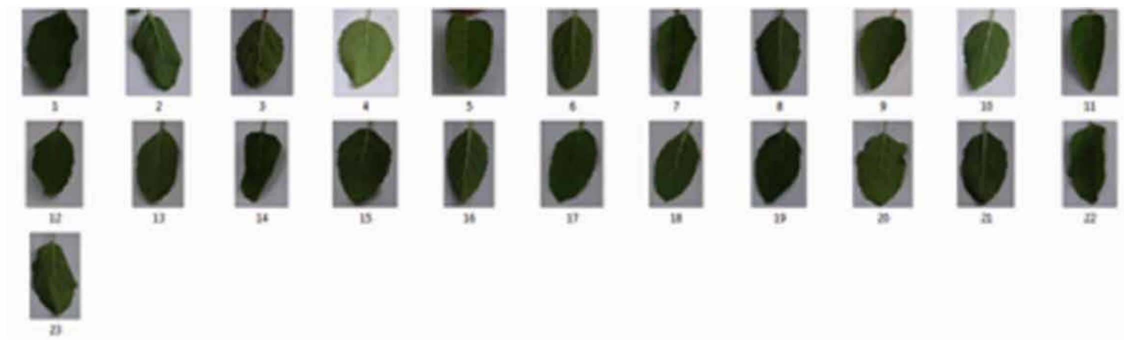
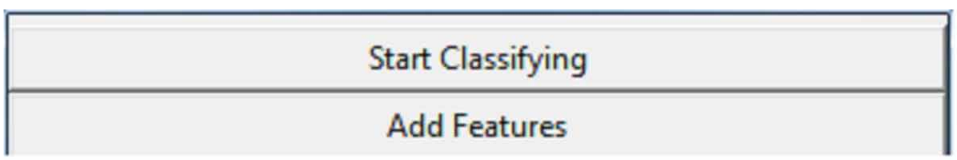


Figure 13. Tool for diseased leaf identification
(Adapted from (Rajesh et. al., 2019))



from the foreground. Whereas in cluster 2, spot diseases are identified on the leaf using the foreground detection. In cluster 3 reveals the accurate spots on the image comparing green pixel and spot pixels et. al., Arbelaitz, Roy, Ashour & Arthur [17, 18, 19, 20].

Figure 15 is the MATLAB figure generated using K-mean algorithm during simulation on the leaf image expressed in Figure 14. Table 1 shows the performance of the proposed method on the dataset considered. We have divided the dataset into training and testing. Three types of experiment are considered. In the first type, 20 samples are used for training and remaining samples are used for testing purpose. In the second type and third type, 30 and 40 samples are considered for training and remaining samples are used for testing purpose respectively. From the table it is quite evident that the proposed KPCA with PNN performs better compared to standard technique.

Feature Extraction

From the input images, the features are to be extracted. To do so instead of choosing the whole set of pixels we can choose only which are necessary and sufficient to describe the whole of the segment. The segmented image is first selected by manual interference. The affected area of the image can be found from calculating the area connecting the components.

First, the connected components with 6 neighbourhood pixels are found. Later the basic region properties of the input binary image are found. The interest here is only with the area. The affected area is found out. The percent area covered in this segment says about the quality of the result. The histogram of an entity or image provides information about the frequency of occurrence of certain value in the whole of the data/image. It is an important tool for frequency analysis. The co-occurrence takes this

Figure 14. Sample images of spot disease identified leaves using k-means clustering method
(Adapted from (Rajesh et. al., 2019))

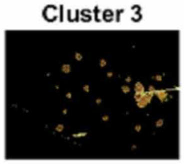
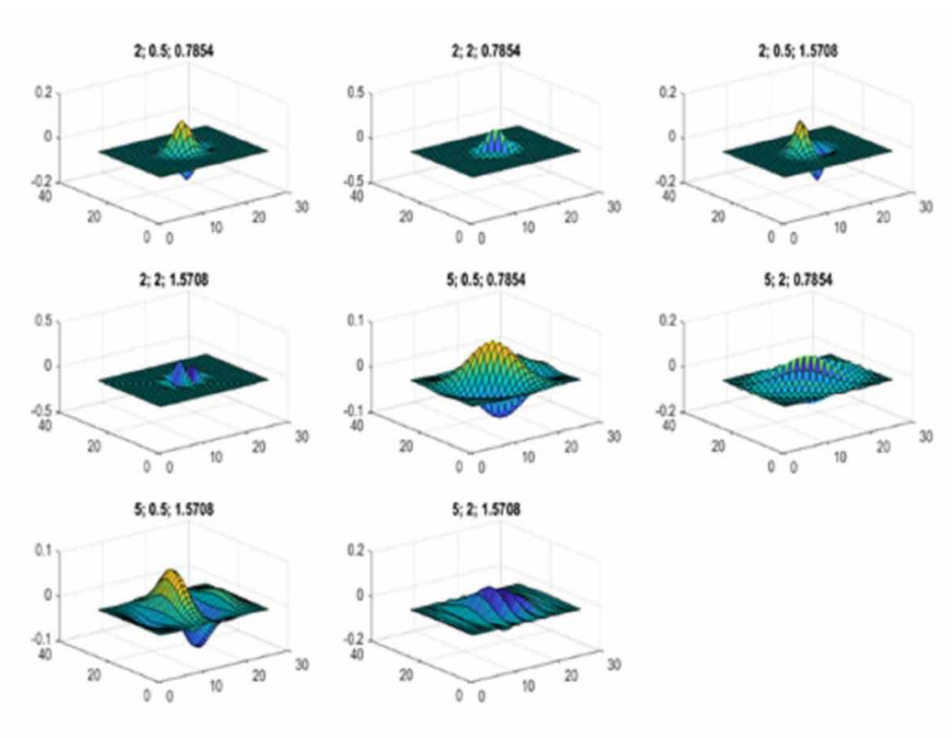
Cluster 1 	Background subtraction to Bifurcate the leaf from the unwanted background Cluster 1 = Background Cluster 2 = Foreground
Cluster 2 	Identification of Spot(disease) in the leaf using the foreground detection Cluster 1: Green (healthy) pixels Cluster 2: Spot(disease) Pixels
Cluster 3 	Accurate result of the detected spots in the leaf Background = Green pixels Foreground = Spot(disease) Pixels

Figure 15. Spot Value Calculation in the disease leaves
(Adapted from (Rajesh et. al., 2019))



analysis to next level wherein the intensity occurrences of two pixels together are noted in the matrix, making the co-occurrence a tremendous tool for analysis (Komali et al., 2015; Nayak et al., 2016; Kumar & Rajesh, 2017).

From grey-co-matrix, the features such as Contrast, Correlation, Energy, Homogeneity are extracted-

Mean: Average or mean value of the array.

Mean is given by

$$\text{Mean}=(1/N)\ Xi \tag{15}$$

where Xi->pixel intensity, N->a total number of pixels of an image.

Standard Deviation: Standard deviation is computed using the below formula:

$$\text{Standard Deviation}= (1/N)\ \Sigma\ (Xi-\mu)\ ½ \tag{16}$$

where μ-> mean.

Entropy: Entropy is a statistical measure of randomness that is used to characterize the texture of the input image. Entropy is defined as

Entropy = -sum (p.*log2 (p)) Where p-> histogram counts. (17)

Variance: Variance is computed using

$$\text{Variance}= (1/N)\ \Sigma\ (Xi-\mu)^{2} \tag{18}$$

Variability is measured using variance. Skewness: The image surface is judged with the Skewness. The same feature set is used for training the SVM as well to identify the class of the input image.

Training

1. Start with images of which classes are known for sure.
2. Find the property set or feature set for each of them and then label suitable.
3. Take the next image as input and find features of this one as new input.
4. Implement the binary SVM to multi class SVM procedure.
5. Train SVM using kernel function of choice. The output will contain the SVM structure and information of support vectors, bias value etc.
6. Find the class of the input image.
7. Depending on the outcome species, the label to the next image is given. Add the features set to the database.
8. Steps 3 to 7 are repeated for all the images that are to be used as a database.
9. Testing procedure consists of steps 3 to 6 of the training procedure. The outcome species is the class of the input image.

10. To find the accuracy of the system or the SVM, in this case, random set of inputs are chosen for training and testing from the database. Two different sets for train and test are generated. The steps for training and testing are same, however, followed by the test is performed.

Classification

The binary classifier which makes use of the hyper-plane which is also called as the decision boundary between two of the classes is called as Support Vector machine (SVM). Some of the problems of pattern recognition like texture classification make use of SVM. Mapping of nonlinear input data to the linear data provides good classification in high dimensional space in SVM. The marginal distance is maximized between different classes by SVM. Different kernels are used to divide the classes. SVM is basically binary classifier which determines the hyper plane in dividing two classes. The boundary is maximized between the hyper plane and the two classes. The samples that are nearest to the margin will be selected in determining the hyper plane are called as support vectors.

Figure 16 shows the concept of support vector machine. Multiclass classification is also possible either by using one-to-one or one-to many. The highest output function will be determined as the winning class. Classification is performed by considering a larger number of support vectors of the training samples. The standard form of SVM was intended for two-class problems. However, in real life situations, it is often necessary to separate more than two classes at the same time. In this Section, we explore how SVM can be extended from binary problems to multi classification problems with k classes where k > 2. There are two approaches, namely the one-against-one approach and the one-against-all approach. In fact, multi-class SVM converts the data set to quite a few binary problems. For example, in one-to-one approach binary SVM is trained for every two classes of data to construct a decision function. Hence there are k (k−1)/2 decision functions for the k-class problem. Suppose k = 15, 105 binary classifiers need to be trained. This suggests large training times. In the classification stage, a voting strategy is used

Figure 16. Linear SVM
(Adapted from (Rajesh et. al., 2019))

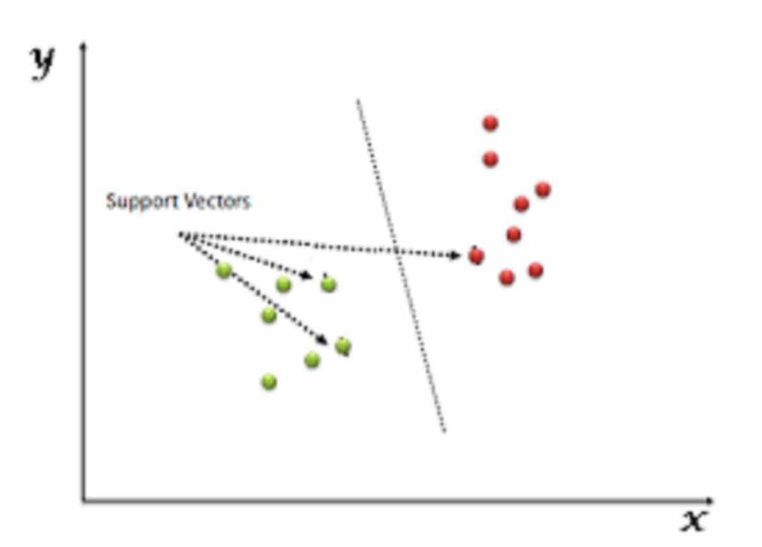
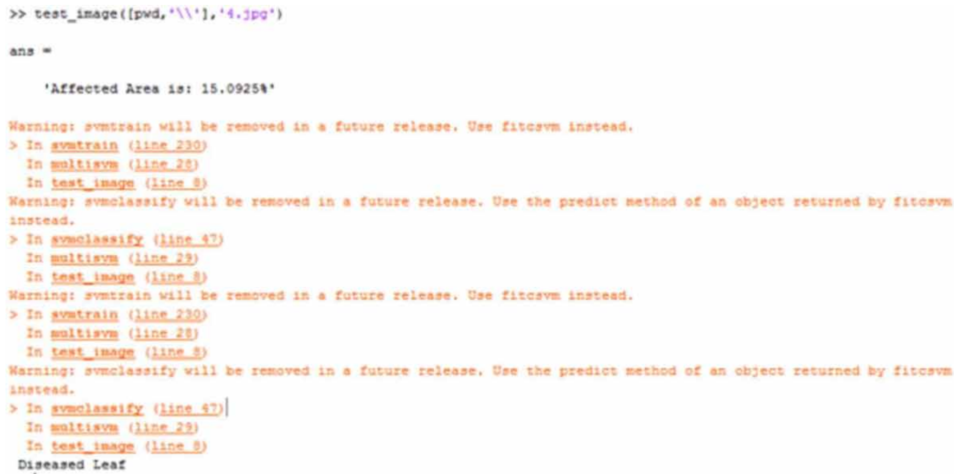


Figure 17. Output as obtained on screen
(Adapted from (Rajesh et. al., 2019))



```
>> test_image([pwd,'\\'],'4.jpg')

ans =

    'Affected Area is: 15.0925%'

Warning: svmtrain will be removed in a future release. Use fitcsvm instead.
> In svmtrain (line 230)
  In multisvm (line 28)
  In test_image (line 8)
Warning: svmclassify will be removed in a future release. Use the predict method of an object returned by fitcsvm
instead.
> In svmclassify (line 47)
  In multisvm (line 29)
  In test_image (line 8)
Warning: svmtrain will be removed in a future release. Use fitcsvm instead.
> In svmtrain (line 230)
  In multisvm (line 28)
  In test_image (line 8)
Warning: svmclassify will be removed in a future release. Use the predict method of an object returned by fitcsvm
instead.
> In svmclassify (line 47)
  In multisvm (line 29)
  In test_image (line 8)
 Diseased Leaf
```

where the testing point is designated to be in a class having the maximum number of votes. The voting approach is called the "Max Wins" strategy. In one-against-all approach, there will be one binary SVM for each of the class to isolate the members of one class from the other class (Rajesh & Aradhya, 2016; Kayyashree & Rajesh, 2018; Guarav, Rajesh & Shaila, 2018; Jena, Rajesh & Patil, 2019).

ANALYSIS AND COMPARISON OF EXPERIMENTAL RESULTS

For the analytic discussion, the K-mean algorithm is implemented and later integrated with PCA using the best or elite solutions form the population in order to develop the ICA-K-mean algorithm. In the simulation, the value of k is kept at the range of 5 to 50 and d is set to 0.1. The comparison among PCA and ICA and has been represented in the Table 2, 3, 4 and 5.

In order to evaluate the performance of the proposed system, we have taken diverse spot disease leaves images from the available sources like plant dataset in the websites. This dataset consists of 100 individual spot disease leave images and each consists of 50 samples. Figure shows the sample spot disease identified leaves.

Dataset is divided into training and testing phases. Three kinds of experimentation are considered. In the first phase, 20 sections are used for training and lasting samples are used for testing persistence. In the second phase and third phase, 30 and 40 samples are measured for training and lasting samples are used for testing persistence correspondingly. From the Table 2 it is pretty clear that the projected KPCA performs better compared to standard technique.

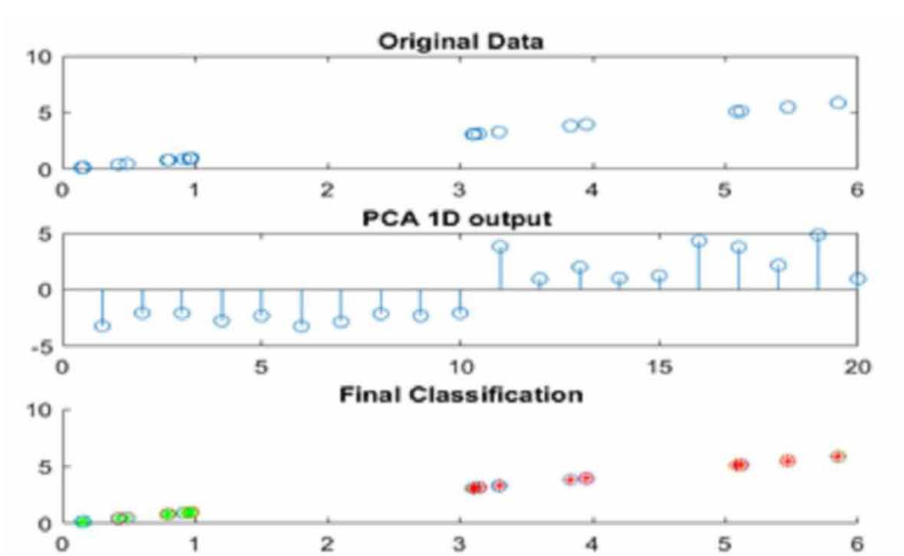
In Figure 19, the original data section shows the whole content of the leaf in terms of clusters, including green and other colored pixels. In the PCA 1D output stage shows the classification of the identical colored clusters obtained from the image. In final classification stage, green dots indicates healthy portion of the leaf and red dots indicates number of spot values (disease) present in the leaf.

Figure 18.

Table 2. Analysis of the spots accuracy in the plants using proposed method

Training: Testing / Methods	20:30	30:20	40:10
PCA	74.6	83.1	89
KPCA	76.5	83.5	91.1

Figure 19. PCA for Spot Value Calculation
(Adapted from (Rajesh et. al., 2019))



TESTING

Test Case 1: Healthy Leaf

In the clustering process, user is expected to input the Region of Interest (cluster number): See Figure 21

Based on Cluster number chosen, three cases are available.

However, irrespective of cluster, expected output is Healthy Leaf.

Test Case 2: Diseased Leaf

In the clustering process, user is expected to input the Region of Interest(cluster number):

Based on Cluster number chosen, three cases are available.

Plant Disease Verification System Using Histogram of Gradients and Kernel PCA

The proposed model contains two phases viz., spot disease identification and spot disease verification. In case of spot disease enrolment, the leaf is enrolled into the database by generating a unique spot disease identification number (ID). During the enrolment phase the leaf shall provide a set of its templates called training leaf templates shall be enrolled into the database across the identification number. In verification stage, a leaf who intends to claim identity shall provide the ID and test template leaf to the system. The test template leaf shall be compared with those of the template leaf that are stored in the database across the ID. If the distance between the test template and the training image is within a predefined threshold, then that leaf shall be declared as genuine leaf else declared as diseased.

Disease Spot Enrolment

Each leaf has to be enrolled into the database prior to verification. Enrolment involves data acquisition, pre-processing, suitable feature extraction and storing the disease plant leaves templates into the database. Along with this for each leaf disease, a unique identification number ID is generated which shall be used during the verification stage.

For each disease leaf a set of ‘N’ leaves are acquired using a suitable acquisition sensor and converted into digital format. Processing on original leaf is time consuming, also requires huge amount of memory and might not be accurate. In order to procedure efficiently, the digital leaf images might be converted into a single vector format by extracting suitable features. Prior to extracting features, the disease leaf has to be pre-processing so that image is good enough to extract to extract the features. In this work we have used following pre-processing techniques;

1. Disease leaf image size normalization: All the images need to be normalized into a standard size such that the leaf yields same feature dimension.
2. As the leaf is captured using digital scanner, the image might contain noise. The following are the reasons for noise in the image, (i) The quality of the photo might not be good which introduces some equally distributed noise (ii) The dust on the scanner introduces some speckle noise. For,

Figure 20. A healthy leaf
(Adapted from (Rajesh et. al., 2019))

Figure 21. Clusters created for healthy leaf in Figure 20
(Adapted from (Rajesh et. al., 2019))

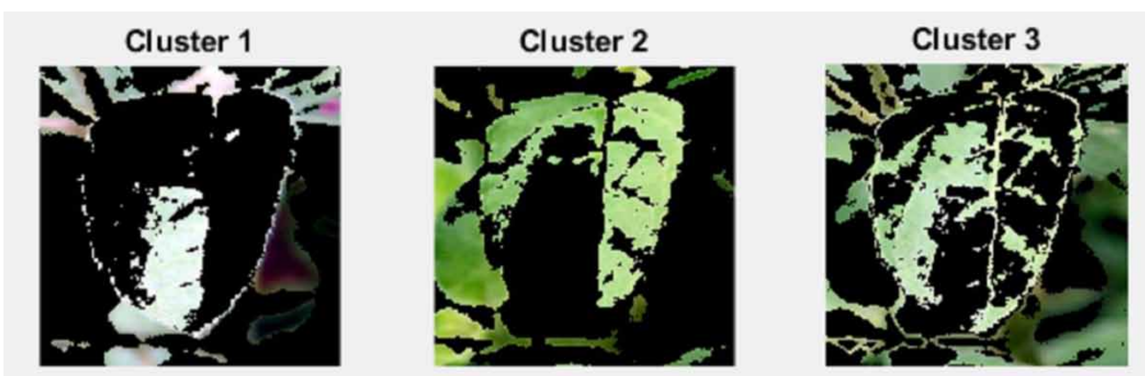


Table 3.

	Input	Actual Output
Case 1	Cluster 1	Diseased Leaf
Case 2	Cluster 2	Healthy Leaf
Case 3	Cluster 3	Healthy Leaf

Figure 22. Clusters shown for the diseased leaf in Figure 20
(Adapted from (Rajesh et. al., 2019))

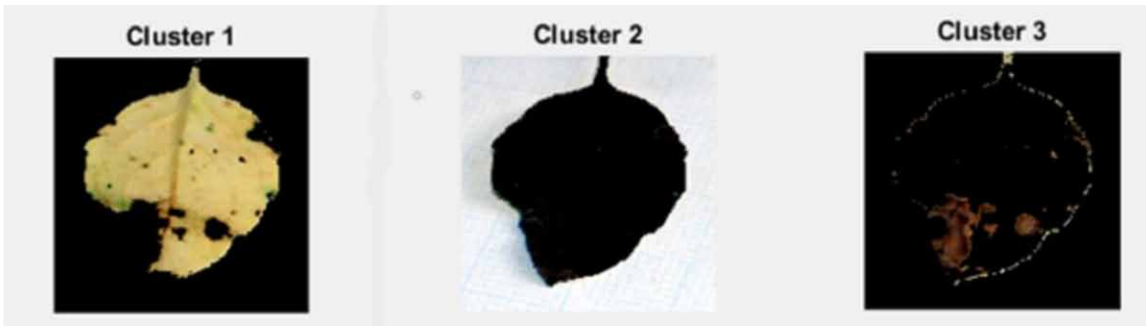


Table 4.

	Input	Actual Output
Case 1	Cluster 1	Diseased Leaf
Case 2	Cluster 2	Diseased Leaf
Case 3	Cluster 3	Diseased Leaf

Expected Output-Diseased Leaf
Note: It is a semi-automatic system as it involves a human loop factor.

first type of noise can be avoided by suitable thresholding technique. Whereas for second type of noise median filter is used.

3. Another important issue in the disease leaf processing is the direction in which the spot is recognized. This effect the estimation of features in obtaining geometric transformation free features. So, we carryout alignment of leafs by rotating the leaf onto a zero degree orientation at centre of mass of leaf.

After pre-processing diseased leafs, a suitable feature extraction is carried on the pre-processed leafs. In this work we have extracted histogram of gradients [HoG]. The reason for selecting HoG features is that, the HoG features are capable to preserve the direction of each pixel with respect the neighbouring pixels which will be unique with respect to the leafs of different disease leafs. HoG can be extracted globally on the entire image or locally. Hence in this model we extract local HoG by the splitting the leaf image into number of blocks. For each block we extract the HoG and we concatenate the histograms to form a feature vector which leads to higher dimension. However, using all features might create curse of dimensionality problem. Hence in this work we recommend to reduce the dimension of the features by applying kernel PCA. Selecting the number of reduced features is a critical issue and hence in this work we choose the dimension empirically. Another issue in using kernel PCA, is choose of the kernel, so we have carried out experimentation using three different kernels namely, linear, Gaussian and Polynomial. A database of training leaf templates of all original leaves are created in such a way that, given a leaf identification number, corresponding training leaf templates can be accessed easily. The entire procedure is termed as disease leaf enrolment into the database.

Disease Leaf Verification

In disease leaf verification stage, a disease leaf who claims spot, is supposed to give, identification number and test leaf template. The first two steps that we have recommended for disease leaf enrolment viz., pre-processing and feature extraction have to be applied on the test leaf template. Given a test leaf template, we obtain the same dimension of the feature vector similar to training set.

The obtained feature vector and the disease leaf identification number is fed to matching model, in which the training leaf templates of identification number given by the disease leaf are fetched and compared with that of the test leaf template. In order to perform comparison, we recommend to use simple minimum classifier between the training leafs and the test leaf. If the minimum distance happens to be less than the predefined threshold, the corresponding disease leaf is declared as a genuine disease leaf else declared as an imposter.

The challenging issue is selection of the suitable threshold, in order to tackle this issue we have used calculation of false acceptance rate (FAR) i.e. the rate at which the number of times the spots are falsely accepted. The false rejection rate (FRR), the rate at which number of times the spots has been rejected falsely, under varying threshold. The point where both FAR and FRR are equal is selected as a corresponding threshold and that point is termed as equal error rate (ERR). In the experimentation section we have provided the performance of the proposed model based on the EER.

$$EER=(FAR+ FRR)/2 \tag{19}$$

The HoG feature can be extracted globally or locally and each one have their own advantages and limitations. In order to study this we have extracted the features in the following ways (i) On the entire leaf image, (ii) splitting the image into 2X2 and 5X5. For each above splitting we have applied kernel PCA with 2 kernels viz., Gabor and linear under varying dimensions from 1 to 2. Below Tables 5-9 shows the results obtained for different image splitting, kernels and dimensions.

ICA shows high FAR rate and low FRR and EER rate should be relatively same for different dimension and different partitions in better performing algorithm. Here ICA shows less difference between FAR and FRR between Gabour kernel and linear kernel.

The implementation time of the proposed algorithm is fewer in SVM in comparison to k-mean algorithm. The following Table 7 illustrates the statistical significance level of difference between K-mean, SVM. It reports the time taken during the implementation of the proposed algorithm. Among the

Table 5. Result Nearest neighbor as classifier where image is divided into 2 grids (Adapted from (Rajesh et. al., 2019))

Classifier = Nearest neighbor classifier Kernel = Gabor	Number of Partition = 2(dividing the image into partition)
Dimension	**EER**
1	1.178
2	1.178

Table 6. Result using ICA as classifier where image is divided into 5 grids (Adapted from (Rajesh et. al., 2019))

Classifier = Nearest neighbor classifier Kernel = Gabor	Number of Partition = 5
Dimension	**EER**
1	1.178
2	1.178

Table 7. Result using SVM as classifier where image is divided into 2 grids (Adapted from (Rajesh et. al., 2019))

Classifier = Nearest neighbor classifier Kernel = Linear Number of Partition = 2	Number of Partition = 2
Dimension	**EER**
1	0.262
2	0.136

Table 8. Result using SVM as classifier where image is divided into 5 grids (Adapted from (Rajesh et. al., 2019))

Classifier = Nearest neighbor classifier Kernel = Linear Number of Partition = 5	Number of Partition = 5
Dimension	**EER**
1	0.262
2	0.136

Table 9. Performance comparison between K-mean, PCA and PCA and ICA algorithm (Adapted from (Rajesh et. al., 2019))

	Feature classification	Identification accuracy
K-mean	96%	90%
KPCA	98%	95%
ICA	98%	98%

different phases of implementation, least time is used for pre-processing as in MATLAB functions like imresize do it easily. It takes more time to make the database for the implementation. We carefully picked images which effectively contributes towards identification and verification of the diseases in plants. Sometimes incorrect angle or lighting blur the important features. It may lead to miss identification of disease in many cases. The selection of appropriate images is one of the crucial phase which contributes the performance of the algorithm.

ICA gives a promising result using EER where equal amount of FAR and FRR results are considered and averaged. These results are subjected to be termed as equal error rate which should show almost same result for different dimension and partitions. With respect to tables 2 to 5 the comparison results using ICA technique with respect to EER shows that the identification and verification of spots in the leaves varies from kernel to kernel. Here Gabor kernel with 1 and 2 partitions give 100% accurate result compared to linear kernel with 1 and 2 partitions. From all the above tables it can be understood that extracting the HoG globally is better than locally and on comparison of different kernels it can be understood that Gaussian kernel performs better than the linear kernel.

CONCLUSION

In this book chapter, an PCA and ICA algorithms are used and compared with SVM and K-means clustering algorithm for identification of spots for plant disease identification. SVM is comparatively new, robust, learning influenced and population based algorithm. It is popular for easy implementation and constraint free parametric settings. Identification and verification of plant diseases by image processing using SVM is a novel approach. The calculated EER is compared with PCA & ICA algorithms. It is found that it provides profound result than only SVM. In future work, the proposed techniques can be tested with standard benchmark real datasets to ensure its robustness and performance quality.

REFERENCES

Al-Hiary, H., Bani-Ahmad, S., Reyalat, M., & Braik, M., & Al-Rahamneh, Z. (2011). Fast and accurate detection and classification of plant diseases. *Machine Learning*, *14*(5).

Aradhya. (2007). Language Independent Skew Estimation Technique Based on Gaussian Mixture Models: A Case Study on South Indian Scripts. *Lecture Notes in Computer Science*.

Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J. M., & Perona, I. (2013). An extensive comparative of cluster validity indices. *Pattern Recognition*, *46*(1), 243–256. doi:10.1016/j.patcog.2012.07.021

Arthur, D., & Vassilvitskii, S. (2006, June). How slow is the k-means method? In *Proceedings of the twenty-second annual symposium on Computational geometry* (pp. 144-153). ACM. 10.1145/1137856.1137880

Ashour, A. S., Samanta, S., Dey, N., Kausar, N., Abdessalemkaraa, W. B., & Hassanien, A. E. (2015). Computed Tomography Image Enhancement Using Cuckoo Search: A Log Transform Based Approach. *Journal of Signal and Information Procedureing*, *6*(03), 244–257. doi:10.4236/jsip.2015.63023

Ashour, A. S., Samanta, S., Dey, N., Kausar, N., Abdessalemkaraa, W. B., & Hassanien, A. E. (2015). Computed Tomography Image Enhancement Using Cuckoo Search: A Log Transform Based Approach. *Journal of Signal and Information Procedureing*, *6*(03), 244–257. doi:10.4236/jsip.2015.63023

Ashourloo, D., Matkan, A. A., Huete, A., Aghighi, H., & Mobasheri, M. R. (2016). Developing an index for detection and identification of disease stages. *IEEE Geoscience and Remote Sensor Letters*, *13*(6), 851–855.

Bonanomi, G., Antignani, V., Capodilupo, M., & Scala, F. (2010). Identifying the characteristics of organic soil amendments that suppress soilborne plant diseases. *Soil Biology & Biochemistry*, *42*(2), 136–144. doi:10.1016/j.soilbio.2009.10.012

Gaurav, Viveka, Rajesh, & Shaila. (2018, February). Automated Number Plate Verification System based on Video Analytics. *International Research Journal of Engineering and Technology*.

Jena, T., Rajesh, T. M., & Patil, M. (2019). Elitist TLBO for Identification and Verification of Plant Diseases. In A. Kulkarni, P. Singh, S. Satapathy, A. Husseinzadeh Kashan, & K. Tai (Eds.), *Socio-cultural Inspired Metaheuristics. Studies in Computational Intelligence* (Vol. 828). Singapore: Springer. doi:10.1007/978-981-13-6569-0_3

Kanungo, D. P., Nayak, J., Naik, B., & Behera, H. S. (2016). Hybrid clustering using elitist teaching learning-based optimization: An improved hybrid approach of PCA. *International Journal of Rough Sets and Data Analysis*, *3*(1), 1–19. doi:10.4018/IJRSDA.2016010101

Kanungo, D. P., Nayak, J., Naik, B., & Behera, H. S. (2016). Hybrid clustering using elitist teaching learning-based optimization: An improved hybrid approach of PCA. *International Journal of Rough Sets and Data Analysis*, *3*(1), 1–19. doi:10.4018/IJRSDA.2016010101

Kavyashree, D., & Rajesh, T. M. (2018, November). Analysis on Text Detection and Extraction from Complex Background Images. *Journal on Pattern Recognition*, 37-43.

Komali, A., Babu, R. V., Kumar, D. S., & Babu, K. G. (2015). An Efficient Content Based Image Retrieval System for Color and Shape Using Optimized K-Means Algorithm. *International Journal of Computer Science and Network Security*, *15*(4), 127.

Korkut, U. B., Göktürk, Ö. B., & Yildiz, O. (2018). *Might. Detection of plant diseases by machine learning. In 2018 26th Signal Procedureing and Communications Applications Conference (SIU)*. IEEE.

Krishna, P. R., & Sao, S. (2016). An improved PCA algorithm to solve profit based unit commitment problem under deregulated environment. *Procedia Technology*, *25*, 652–659. doi:10.1016/j.protcy.2016.08.157

Kumar, A. (2017, April). A Moving Object Recognition using Video Analytics. *International Journal for Scientific Research & Development*.

Nayak, J., Naik, B., Kanungo, D. P., & Behera, H. S. (2016). *A hybrid elicit teaching learning based optimization with fuzzy c-means (ICA-FCM) algorithm for data clustering*. Ain Shams Engineering Journal.

Rajesh, T. M. & Aradhya. (2015). An Application of GMM in signature skew detection. Journal of pattern Recognition.

Rajesh, T. M., & Aradhya, V. N. (2016, September). ICA and Neural Networks for Kannada Signature Identification. *International Journal of Latest Trends in Engineering and Technology*.

Rao, R., & Patel, V. (2012). An elitist teaching-learning-based optimization algorithm for solving complex constrained optimization problems. *International Journal of IndustrialEngineering Computations*, *3*(4), 535–560. doi:10.5267/j.ijiec.2012.03.007

Rao, R. V., & Patel, V. (2013). An improved teaching-learning-based optimization algorithm for solving unconstrained optimization problems. *Scientia Iranica*, *20*(3), 710–720.

Roy, K., Mitra, I., Kar, S., Ojha, P. K., Das, R. N., & Kabir, H. (2012). Comparative studies on some metrics for external validation of QSPR models. *Journal of Chemical Information and Modeling*, *52*(2), 396–408. doi:10.1021/ci200520g PMID:22201416

Sankaran, S., Mishra, A., Ehsani, R., & Davis, C. (2010). A review of advanced techniques for detecting plant diseases. *Computers and Electronics in Agriculture*, *72*(1), 1–13. doi:10.1016/j.compag.2010.02.007

Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., & Stefanovic, D. (2016). Deep neural networks based recognition of plant diseases by leaf image classification. *Computational Intelligence and Neuroscience*, *2016*, 1–11. doi:10.1155/2016/3289801 PMID:27418923

Tichkule, S. K., & Gawali, D. H. 2016, November. Plant diseases detection using image procedureing techniques. In *Green Engineering and Technologies (IC-GET), 2016 Online International Conference on* (pp. 1-6). IEEE.

Wu, H., Zhou, Y., Luo, Q., & Basset, M. A. (2016). Training Feedforward Neural Networks Using Symbiotic Organisms Search Algorithm. *Computational Intelligence and Neuroscience*, *2016*, 1–14. doi:10.1155/2016/9063065 PMID:28105044

Chapter 7
Applications of Data Mining Techniques in Smart Farming for Sustainable Agriculture

Anandhavalli Muniasamy
https://orcid.org/0000-0001-8940-3954
King Khalid University, Saudi Arabia

ABSTRACT

Smart farming is a development that highlights the use of technologies such as the internet of things, cloud computing, machine learning, and artificial intelligence in the farm management cycle. For sustainable agriculture to adapt the ongoing change in climate and social structure is a major challenge for scientists and researchers. The approach needs information from various sources and its use in the relevant field, which lead to a growing interest in knowledge discovery from large data. Data mining techniques provide effective solutions for this problem as it supports the automation of extracting significant data to obtain knowledge and trends, the elimination of manual tasks, easier data extraction directly from electronic sources, and transfer to secure electronic system of documentation, which will increase the agriculture productions from same limited resources. In a nutshell, the aim of this chapter is to gain insight into the applications of data mining techniques in smart farming, which direction to employ sustainable agriculture and identify the challenges to be addressed.

INTRODUCTION

Smart farming (SF) involves farming management concept using modern technology and the integration of information and communication technologies (ICT) into machinery, equipment, and sensors for use in agricultural production systems to increase the extent and value of agricultural products. The smart farming system is an autonomous & sophisticated mechanism, which will aid in the growth of agriculture yield by applying hi-tech agriculture techniques without human intervention. Agricultural practices are expected to advance the leading notion of "smart farming" (Tyagi, 2016; Babinet, Gilles et al., 2015) as it is being supported by biotechnology (Rahman, et al., 2013) and emerging digital technologies such

DOI: 10.4018/978-1-5225-9632-5.ch007

as cloud computing (Hashem, et al., 2015), artificial intelligence (AI) & robots, Internet of Things (IoT) (Weber & Weber, 2010), and remote sensing (Bastiaanssen, et al., 2000).

Traditional farming practices are resource intensive in terms of capital, land, water, and fossil fuel use. But it threatens future food production by reducing biodiversity and contributing to environmental degradation and climate change which lower yields. Modern agriculture focuses on the high degree of mechanization, large scale, use science and technology, high capital and equipment, and use biotechnology to produce organic pollution-free food.

Farmers in the twenty-first century have access to Global Positioning System (GPS), soil scanning, data management, and IoT technologies. Agricultural enterprises have started collecting large amounts of data that includes soil and crop properties, which enables higher operational efficiency also. So, growth in this data size requires an automated method to extract necessary data.

Data mining is the process of finding correlations or patterns among the features in large relational databases. The overall goal of the data mining process is to extract information from a data set to extract previously unknown, interesting patterns such as groups of data records (cluster analysis), unusual records (anomaly detection) and transform it into an understandable structure for further use, which is used for all data mapping & processing. The technology of data mining is narrowly connected to data storage and is intertwined with the database management system. By applying data mining techniques, it is possible to extract useful knowledge, find patterns and trends in huge data.

Smart farming system covers main issues of farming such as water required for the crop at a particular stage, fertilizer to be used according to the micro-nutrients as well as macronutrients present in the soil, and the pesticides to be used depend on various environmental factors such as humidity, pollution (air, soil and water) etc.

Smart farming technologies require more and more professional skills to analyze agricultural data; extracting useful information has become the question of great importance. The implications of these situations highlight the need to consider the applications of data mining techniques to overcome the various problems faced by farmers in agriculture to grow good yield with SF for sustainable agriculture.

Since the growth in agriculture data is huge (Miller, McCarthy & Zakzeski, 2009), information is often hidden, so data mining techniques are used for their detection. Data mining techniques can be useful for finding the schemes of marketing in data, which are valuable and interesting for crop management (Mucherino & Rub, 2011). Both data mining and smart farming are relatively trending concepts, so it is expected to have knowledge about these applications and their implications for research and development be not widely spread.

Agriculture practices have developed from the mule and plow into the high-tech business. Smart agriculture is abundant with diverse information, which conditions the necessity to use data mining. The basic idea is when the same data analyzed in different the context, there will a generation of new context-based information and knowledge which agricultural management uses for improving its decisions. Data mining techniques have made statistics enhanced by the quality of data from collection to evaluation. By ease of use and the possibility of presenting complex results in a simple fashion, data mining has shown to be fertile ground for future innovation in the field of agricultural statistics (Miller et al., 2009).

This chapter mainly focuses on how smart farming can be achieved with descriptive and predictive information as support to decision making with the help of data mining techniques; insights the concepts of smart system, smart farming, precision agriculture, sustainable agriculture, traditional agriculture practices with its pros and cons; describes the smart farming framework for data mining applications and the proposed conceptual framework of data mining applications towards smart farming; discussing the data mining issues in SF applications; provides solutions and recommendations based on previous data mining applications in SF research. Finally, the chapter concludes with a summarization of the study, a balanced assessment of the contribution of data mining applications in smart farming, and a roadmap for future directions.

BACKGROUND AND LITERATURE REVIEW

The smart farming system is an autonomous & sophisticated mechanism, which will aid in the growth of agriculture yield by applying hi-tech agriculture techniques without human intervention. The literature on smart farming is recent (Dieisson, 2018). The concept and terms associated with SF have not reached a consensus in the scientific literature (S. Wolfert 2017). The basis for advancement in SF involves a combination of internet technologies and future-oriented technologies with smart objects (M. Brettel 2014; H. Lasi, 2014; Y. Liao, 2017; A. D. Maynard, 2015); however, there is no still established concept for these technologies in agriculture (S. Wolfert, 2017).

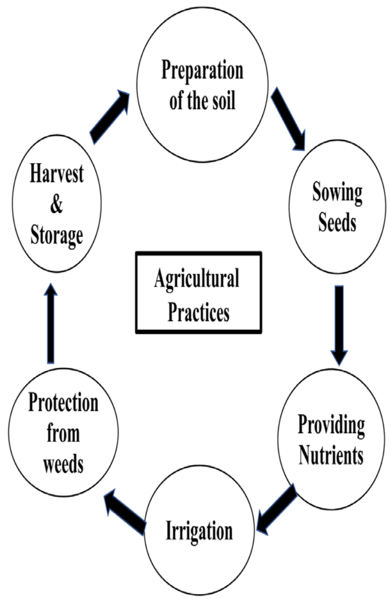
Agricultural Practices

Agriculture or Farming is the cultivation of animals, plants, and other life forms for food and other products used to sustain life. Farming of domesticated species created food surpluses that nurtured the rise of sedentary human civilization.

Agricultural practices mean the steps involved in agriculture which can be seen generally been done by farmers. Agricultural practices such as preparation of soil, sowing, adding manure and fertilizers, irrigation, protection from weeds, harvesting, storage have made great strides in the past century.

- **Preparation of Soil:** The soil needs to be prepared prior to planting, usually by tillage or chemical "burn-down" to kill the weeds in the seedbed that would compete with the cultivated plants for water and nutrients.
- **Sowing:** It is the process of planting seeds, which involves seed soaking, seed cleaning, and seed washing.
- **Use of Manures and Fertilizers:** The nutrient supplying sources are manures and fertilizers, which are required by the plant for its growth and development which are absorbed through the soil. It is one of the important factors which help in increasing the crop yield and to maintain soil fertility.
- **Irrigation:** It is the process of an artificial application of water to the land or soil, which is used to assist in the growing of agricultural crops in dry areas and during periods of inadequate rainfall.
- **Protection From Weeds** or wild plant growing with cultivated plants is important to increase the production rate.
- **Harvest and Storage:** Harvesting crops involves getting the crop out of the field and transported to market. Machines like tractors, harvesters, grain carts, etc. are used to harvest the crops and the harvested grain will be stored to prevent grain quality loss from weather, wind and moisture; birds and insects, and microorganisms.

Figure 1. Agricultural practices



Issues in Current Agricultural Practices

Over the last few decades, many books and articles have been published that draw attention to farmer experimentation and local innovation in agricultural practices (literature includes ILEIA, 2000; Saad, 2002; Reijntjes & Waters-Bayer, 2001). However, farmers mostly adopt traditional methods resulting in low productivity. Traditional farming involves monocropping, synthetic fertilizer, and green revolution technologies. It mainly relies on manpower and experience, in which the production tools are relatively backward and small in scale, limited investment capital equipment and synthetic fertilizers are used in large quantities.

Traditional farming practices are resource intensive in terms of capital, land, water, and fossil fuel use. But it threatens future food production by reducing biodiversity and contributing to environmental degradation and climate change which lower yields. (Challa, 2013) cited that increasing agricultural productivity is critical to meet expected rising demand and, as such, it is instructive to examine recent performance in cases of modern agricultural technologies. (Loevinsohn et al. 2013) concluded, the most common areas of technology development and promotion for crops include new varieties and management regimes; soil as well as soil fertility management; weed and pest management; irrigation and water management. From the study (Margaret Mwangi, 2015), technological, economic, institutional factors and human-specific factors are found to be the determinants of agricultural technology adoption. Modern agriculture focuses on the high degree of mechanization, large scale, use science and technology, high capital and equipment, and use biotechnology to produce organic pollution-free food.

By virtue of improved input/output relationships, new technology tends to raise output and reduces the average cost of production which in turn results in substantial gains in farm income (Challa, 2013). It is therefore important that for any new technology to be introduced to farmers, they should be involved in its evaluation to find its suitability to their circumstances (Karugia et al., 2004).

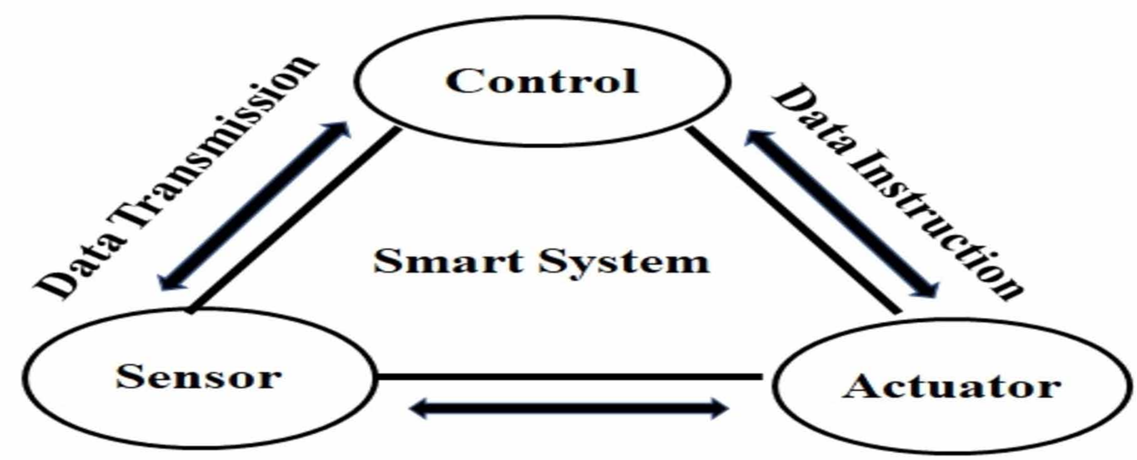
Understanding the factors that influence or hinder adoption of agricultural technology is essential in planning and executing technology-related programs for meeting the challenges of food production in developing countries. Therefore, to enhance technology adoption by farmers, it's important for policy-makers and developers of new technology to understand farmers need as well as their ability to adopt technology in order to come up with technology that will suit them.

Smart System

Smart systems (Figure 2) performing smart actions by integrating functions such as sensing, actuation, and control in order to describe and analyze a situation and make decisions based on the available data in a predictive or adaptive manner. It is the confluence of various components like data acquisition, data transmission, command & control unit, data instruction unit and action devices.

Data acquisition unit aims to collect the required raw data needed for appropriate sensing and monitoring of the structure. Data transmission forward the raw data to the central command or control unit. The role of command & control unit is to manage the whole system by analyzing the data and reaching the appropriate decision and determining the actions required. Data instructions part transfer the decisions and respective instructions to the members of the whole system. Action devices are used to take actions by triggering the controlling unit or devices.

Figure 2. Smart system



Smart Farming

The smart farm also called connected farm can support various devices from diverse agricultural device manufacturers, embedded with ICT and IoT systems, and provide intelligent agricultural services based on shared expert knowledge. For example, people having little experience in farming grow plants or crops for profits, prevent infectious plant disease and take precautionary action to proactively isolating that infectious one from others. (Ryu 2015).

Sundmaeker et al. (2016) defined SF as the phenomenon derived from the rapid developments in the IoT and Cloud Computing. SF allows a large volume of data and information generated with the incorporation of ICT into machinery, equipment, and sensors in agricultural production systems. SF goes beyond precision agriculture (Lokers, et al., 2016), by enhancing management tasks based on location and data, enhanced by context (Kamilaris, et al., 2016), situation and location awareness (Karmas, et al., 2016), initiated by real-time events (Wolfert et al., 2014). SF relies on data transmission and focuses on data in remote storage systems to enable the analysis of various farm data for decision-making.

Smart Farming System

The SF system is the confluence of various systems involves smart sensor & smart mapping, smart equipment monitoring, smart crop monitoring, smart climate & forecast monitoring, smart logistic, stats and warehousing and smart predictive for corps management. The basic view of the SF system is shown in Figure 3.

Components of Smart Farming

SF concept is made up of several, equally important components. Each of these components (Figure 4) is having its own importance and smart farming as a whole is becoming more advanced, as a result.

The components are as follows:

- Sensing Technologies
- Software applications
- Communication systems
- Telematics positioning technologies
- Hardware and software systems
- Data Analytics solutions
- Mobile applications can now be used by farmers to remotely track and manage yields, costs, and other important farm metrics. In a number of cases, mobile operators are likely to provide these capabilities to LPWA users as cloud services that can be easily accessed via application programming interfaces (APIs). Additionally, operators can manage LPWA networks in the same way they manage existing networks. the system with application equipment to apply inputs at a precise time or location to achieve site-specific application rates of inputs.
- Sensing technologies (on-field sensors) have proved mighty useful. Smart farming solutions work through sensors. Farmers can monitor various conditions like soil moisture, water level, light,

Figure 3. Smart farming system

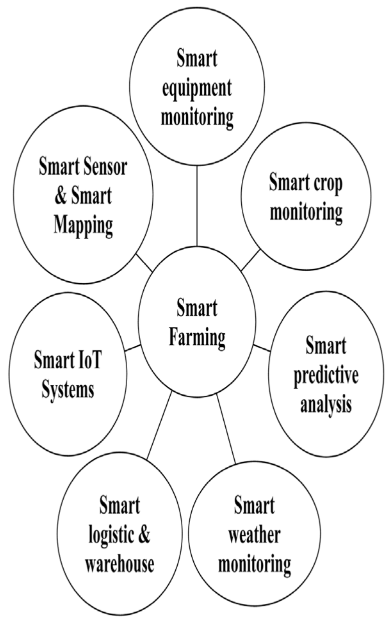
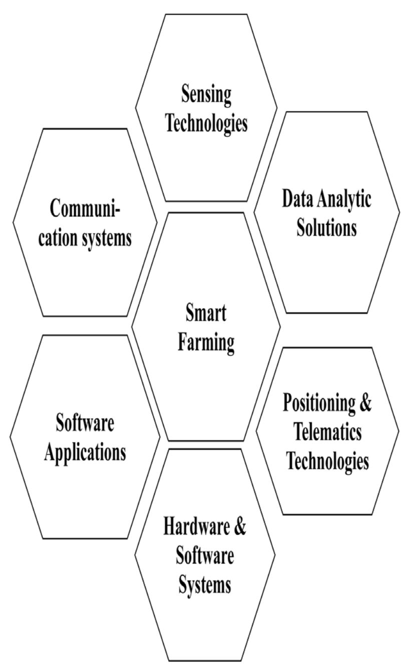


Figure 4. Components of smart farming



humidity, obstacles, and motion from anywhere by combining sensors, motion detectors, button camera, and wearable devices. The IoT-based smart farming automates the irrigation system and is highly efficient as compared to the conventional operations.

- Hardware tools and software solutions: have increased in popularity. Major IoT applications for farming are farm vehicle tracking, livestock monitoring, water level monitoring, storage, and soil monitoring, etc.
- Software applications: It is easy to follow the trends in organic farming, family farming, etc. With the help of IoT-based farming applications. The smart farming apps serve many objectives ranging from managing irrigation and fertilization to providing support to local agronomy as per different seasons.
- Smart positioning technologies (GPS) have done their bit towards making agricultural practices smarter. Telematics (i.e., the transmission of information over long ranges) has been a key component of smart farming as well, as have been the advanced data analytics tools and platforms.
- Communication technology – via the cellular platform – cannot be overemphasized either. From telecom's perspective, providing mainly connectivity services has the immense potential to heavily influence multiple facets across the value chain. In the future, mobile operators can provide three categories of services in smart agriculture for increased market share. Connectivity – Most telecom operators around the world are offering connectivity services, but these services only represent a highly limited proportion of the entire smart agriculture market. Vertical Integration – Telecom operators can offer end to end solutions in smart agriculture and not just connectivity services. This will certainly lead to an exponential increase in the market share of mobile operators. However, vertical integration demands significantly higher investment be made in this area before further measures and benefits can be seen. The smart farm, embedded with IoT systems, could be called a connected farm, which can support a wide range of devices from diverse agricultural device manufacturers. Also, connected farms could provide more intelligent agricultural services based on shared expert knowledge. For example, people having even little experience on farming could grow plants or crops for profits. Infectious disease prevention could also be another benefit of developing the connected farms by detecting influenza virus in a specific pig farm, and proactively isolating that one from others.

Mobile applications are useful for the farmers to remotely track and manage yields, costs, and other important farm metrics as cloud services via application programming interfaces (APIs). Sensing technologies involve on-field sensors which help the farmers to monitor various conditions like soil moisture, water level, light, humidity, etc. and automates the irrigation system by combining sensors, motion detectors, button camera, and wearable devices. Hardware tools and software solutions for farming in IoT applications are important for farm vehicle tracking, livestock monitoring, water level monitoring, storage, and soil monitoring, etc. and software applications serve many purposes ranging from managing irrigation and fertilization to providing support to local agronomy as per different seasons. Smart positioning technologies with GPS are making agricultural practices smarter. Telematics, the transmission of information over long ranges, has been a key component of SF as well. Communication technology, via the cellular platform, provide various services in smart agriculture for increased market share.

Applications of Smart Farming

With the advent of IoT, SF applications are making their mark with the promise to revolutionize the agriculture sector. Some of the important applications are plant and soil monitoring, using drones and AI, remote weather monitoring, precision farming, smart irrigation, yield prediction, smart irrigation, smart greenhouses, and farming management system.

Plant and Soil Monitoring

Soil monitoring system monitors physical, chemical, and biological properties such as texture, water-holding capacity, and the absorption rate and assists farmers in tracking and improving the quality of soil for fertilization and chemical composition monitoring, water usage monitoring for optimal plant growth, weather condition reporting and avoid degradation.

Use of Drones and AI for Farming Practices

In the current scenario, drone technology offers the real-time data and revamps the agriculture industry to perform a variety of tasks and improve various farming practices such as crop health assessment, crop monitoring, crop spraying, planting and land analysis, agriculture photography for site-specific development, sprinkling pesticides, capturing farm photos using GPS technology giving inputs about water and fertilizers level. According to (Grand View Research), the market size for these agricultural drones to exceed USD 3770 million by 2024. These drones are easily controlled by any person either manually or by using on-ground IoT sensors. The drones sent collected data to servers or systems from where the end user can get them. With the help of self-driving tractors, the farmers can control their far, remotely and save the labor costs, which enable them to create the best strategy & planning based on real-time data collection and processing.

Remote Weather Monitoring

With the help of umpteen number of customizable sensors, the farmers can monitor their crops and the weather conditions from remote places to measure the key conditions like temperature, rainfall, humidity, moisture, and chemical composition accurately to get more crop throughout the year and increase productivity without putting additional efforts.

Precision Farming

Precision farming is an approach to farm management for ensuring profitability and sustainability while protecting the environment. It uses IoT and ICT technologies to optimize returns and ensure the preservation of resources and IoT sensors are used to collect data on weather, soil humidity, air quality, and crop maturity; enabling them to make better decisions. The sensor devices and collected data help farmers to better utilize water resources.

Smart Greenhouses

The smart greenhouse allows farmers to cultivate crops with minimal human intervention by using IoT sensors to intelligently monitor & control the internal climatic conditions like temperature, humidity, luminosity, and soil moisture; saving manual labor costs, production losses, and energy.

Smart Irrigation

The conservation of existing water resources by employing sustainable and efficient irrigation systems to minimize water losses. Precise requirements for water are calculated by IoT-based smart irrigation system which measures various parameters such as soil moisture, humidity, light intensity, and temperature to maintain higher irrigation efficiency.

Yield Monitoring

Yield monitoring assists the farmers by monitoring various features corresponding to agricultural yields such as moisture content, grain mass flow, and the total quantity of harvested grain and offers real-time information to farmers to facilitate decision making and reduce operational costs with enhanced productivity.

Farm Management System (FMS)

FMSs provides management and information collection to farmers and other stakeholders for leveraging various sensors and tracking devices. The collected information is then stored and analyzed for conducting complex decision-making tasks. The Kaa IoT Platform is an enterprise-grade IoT enablement technology that connects together different sensors, connected devices, and farming facilities. Kaa (Kaa Project) streamlines the development of smart farming systems and provides maximum flexibility for custom-tailored architecture design.

Sustainable Agriculture

Every person involved in the food system like retailers, consumers, growers, distributors, and waste managers play a role in ensuring a sustainable agricultural system. It focuses on food and fabric needs of society in the present without compromising the facility of future generations to meet their own needs. Sustainable agriculture (Senanayake, 1991) is very relevant and directly linked to smart farming (Bongiovanni & Lowenberg-DeBoer, 2004), as it enhances the environmental quality and resource base in which agriculture depends, providing basic human food needs (Pretty, 2008). Sustainable agriculture integrates a healthy environment, economic profitability, and social and economic equity into agricultural tasks.

Precision Agriculture

Precision agriculture is about real-time data gathering, processing, and analysis, as well as automation technologies on the farming procedures, allowing improvement of the overall farming operations and management, and more informed decision making by the farmers. It enables the vision of smart farming with sustainable agriculture to help farmers to produce more food and conserves soil for sustainable food production, results in stable food supply (Precision agriculture, 2015). Precision agriculture is a data-driven approach to agriculture, which is strongly connected with a various data mining method (Ruß & Brenning, 2010).

Data Mining

Data mining is the process of extracting important and useful information from large sets of data (Abello et al. 2002; Klosgen and Zytkow 2002; Pardalos et al. 2007, 2008). Data mining is the process of extracting important and useful information from large sets of data (Abello et al. 2002; Klosgen and Zytkow 2002; Pardalos et al. 2007, 2008).

Data mining is the process of investigating large pre-existing databases in order to generate new information (Pardalos et al. 2007, 2008; Abello et al. 2002; Klosgen & Zytkow 2002). (Jiawei, Micheline, & Jian 2012) defined data mining as a process of using large data sets to infer important hidden knowledge. Data mining assists enterprises to predict future trends by sorting through large data sets to identify patterns and establish relationships to solve problems through data analysis. Knowledge Discovery in Database (KDD) is the process of discovering the knowledge from structured (relational databases, XML) and unstructured (text, documents, images) data (C. Poultney, 2006). Data mining is a particular step in this KDD process (U. M. Fayyad, 1996; S. Zhang 2003) for identifying valid, novel, potential, useful patterns from the mass of data (G. Hinton, 2006), the results can be further utilized in an automated decision support system.

Data mining is a cyclic process (Figure 5) and each phase has its own importance.

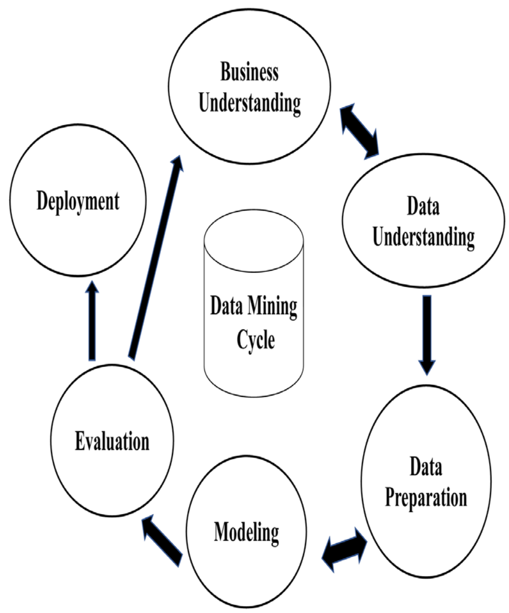
- Define the problem with identifying business goals and data mining goals.
- Identify the required data by collecting it from various sources and understanding of the data.
- Prepare and pre-process the selected require data as necessary.
- Model the data using data mining algorithms and build predictive models.
- Train and test the developed models with sample datasets and iterate the process until getting expected results.
- Verify final models, prepare visualizations for achieved results and deploy the model.

For predicting future trends of agriculture processes, data mining is used for examining data by summarizing in a different perspective and converting it into beneficial information in large datasets. Data mining has no restriction for analyzing the type of data and used for any kind of data including database data, data warehouse data or transactional data.

Data Mining Techniques

Data mining techniques draw the importance of its wide application in agriculture and allied sectors. DM tasks are categorized as descriptive data mining and predictive data mining. Descriptive methods characterize the general properties of the data in the database while predictive methods used to predict unknown or future values based on patterns determined from known results. The main data mining techniques include Classification, Clustering, Association rules, and Regression methods. (Mucherino, Papajorgji & Pardalos, 2009) discussed various data mining techniques used for solving the different ag-

Figure 5. Data mining cycle



ricultural problem as shown in Figure 6. In most of the agriculture tasks, predictive data mining approach is used. Predictive data mining technique used to predict future crop cultivation, weather forecasting, use of pesticides and fertilizers, generation of revenue and so on.

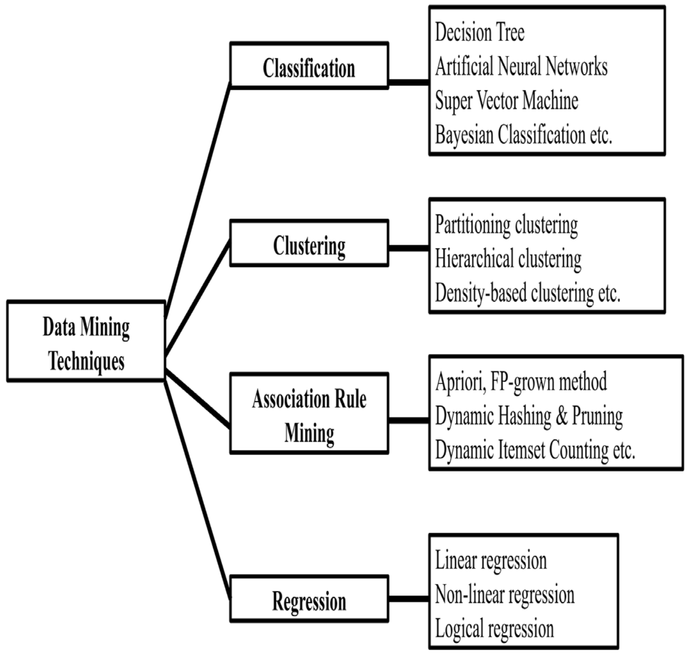
Classification

Classification is a technique which forms the common stepping stone for various recursive algorithms and methods designed for data mining. The classification algorithm is to divide various data-sets or database which have enough data into different types of sets so that they can be analyzed using a single step exhaustive model. Classification method makes use of statistical techniques like decision trees, linear programming, neural network and statistic-oriented approach (Aashay Kumar, 2016), and Rule Based Classifiers, Bayesian Networks, Decision Tree, Nearest Neighbour, Artificial Neural Network, Support Vector Machine (SVM), Rough Sets, Fuzzy Logic, Genetic Algorithms Beniwal, S., & Arora, J. 2012).

Clustering

Clustering is a descriptive data mining technique based on unsupervised classification, called clustering or exploratory data analysis, no labeled data are available (B. Everitt, 2001; A. Jain, 1988). Clustering analyzes data objects without consulting class labels. The objects are clustered or grouped based on the principle of maximizing the intra-class similarity and minimizing the interclass similarity (Jiawei,

Figure 6. Data mining techniques



Micheline, & Jian, 2012). Clustering method identifies dense and sparse regions in object space and can find overall division pattern among data attributes. The broadly used clustering methods are ds are Hierarchical Methods, Partitioning Methods, Density-based Methods, Model-based Clustering Methods, Grid-based Methods and Soft-computing Methods fuzzy, neural network based, Squared Error Based Clustering, Clustering graph and network data (Han, J, Kamber, M., & Pei, J. 2006; Xu, R & Wunsch, D, 2005). Anomaly detection is performed to see which record sticks out from the rest (Chandola, Banerjee, & Vipin 2009).

Association Rules

Association rule mining is a descriptive data mining task which involves determining patterns, or associations, between the elements in the dataset based on a relationship of a particular item on other items in the same transaction (Aashay Kumar, 2016; Neha Patel, 2016). Associations are presented in the form of rules or implications. Mainly this method focuses on finding relationships between the different items in a transactional database. The different association rule mining algorithm are Apriori algorithm, Partition, Dynamic Hashing, and Pruning, Dynamic Itemset Counting, FP Growth, Eclat & Declat, MaxEclat, etc. (Zhang, N., Wang, M., & Wang, N. 2002).

Regression

Regression analysis is a predictive modeling technique which gives the relation between the independent variable and dependent variable and to fit a formula to a dataset. The variable that is been predicted is the dependent variable and the variable which are predicted is used to predict the values of the dependent variable is called the independent variable. The process for estimating the relationships among variables is known as regression analysis which is used to understand how the typical value of the dependent variable can be changed (J. Han, 2012; Aashay 2016). The methods for prediction are linear regression and nonlinear regression.

Data Mining Applications in Various Domain

Data mining has wide application domain almost in every industry where the data is generated that's why data mining is considered one of the most important frontiers in database and information systems. (Buttle, 2009) claim that data from data mining can be used primarily for division of market and customers into segments; Identification of valuable clients and potential customers in the future; Investigation of causes in customer behavior; Defining different prices for individual customer segments; Identification of poor payers; Creating customer profiles that the organization desires to acquire and keep; Identifying successful tactics for keeping and acquiring customers. Data mining is widely used for banking industry for financial data analysis (Kamesh, V., et al., 2019; Nathan Asha, 2016; Madan Lal Bhasin, 2006, S.S.Kaptan, 2002), marketing (Manik Sharma et al.,2018; B. Desai & Anita Desai, 2004; Shaw, Michael J., et al., 2001), retail industry(Berry et al. 2004; Ahmed, S.R., 2004), telecommunication industry (Njenga, 2012; Rayan Masoud 2016; Weiss, Gary. 2009), biological & medical data analysis (Emre et al., 2019; Sharma Singh 2017; Arun, 2004).), intrusion detection (Manoj 2011; Huy & Deokjai, 2008), education (Bakhshinategh, 2018; S. Parack 2012; B. Namratha 2016) and other scientific applications (Grossman, 2001; Saima et al. 2018).

Existing Systems for Data Mining Application for Smart Forming

Milos Brajovic (2015) proposed the irrigation system software which includes irrigation system hardware (sensor) and software (desktop application) with Iris scheduling software. It is a freeware application, which provides an expert system-based irrigation advice using automatically collected or manually inserted data. Various kinds of sensors are used for different purpose such as to monitor air humidity radiation, soil temperature, soil moisture, air temperature, etc. Other sensors are used for monitoring plants their growth rate, chemical exchange, the response on irrigation. This software uses an expert system and artificial intelligence algorithms for interpretation of data. On the basis of this data, it provides a real-time decision about current irrigation time as well as about the future. This system has complex software that gets information from different sources such as weather station, sensors satellite information, and data from the Internet and from the database available.

The mobile-based irrigation system (Ankit Patil 2016) use Arduino UNO (open source prototyping platform depend on user friendly hardware and software) which can be customized according to user need. Arduino UNO board consists of microcontroller used to control motor. It is programmed in such a way that it can sense moisture level in soil with hydrometer sensor and inform Arduino UNO. If the moisture level is below the threshold for a particular crop, then notification of water required will be sent on farmers mobile App from the Arduino UNO. Both of these techniques are costly, as it requires huge hardware and also if the sensor stops working then the whole system will go down or not give an accurate result. In mobile application some of the factors are considered for water required calculation, so, we need a more effective mechanism that gives result even when hardware fails. With the help of data mining smart farming system, we can calculate the water required by finding correlations or patterns among dozens of fields in large relational databases.

Muzammil Hussain (2015) aimed to develop a wireless three levels controlled smart irrigation system to provide an automatic irrigation system for the plants which help in saving water and money. The main objective is to apply the system for improvement of the health of the soil and hence the plant via multiple sensors. In the last few years, remotely monitored embedded system for irrigation purposes have become a new necessity for the farmer to save his energy, time and money. But this system covers only irrigation module.

Crop fertilization recommendation system (Z. Ren & X. Lu,2012) recommends the required amount of nutrients given during the crop growth, choosing suitable fertilizers, and arranging fertilization time. It is a specific computer program to control local entities with the help of the rules form introduced to the system under a certain strategy and produce an applicable recommendation. Whether it can be used widely or not, the drawback is that the models or the parameters in the system can be customized easily according to the local agricultural production practices only.

Methodology

The smart farming framework with data mining techniques (Figure 7) involves various tasks to analyze the selected set of agriculture data where the process involves of preprocessing with cleaning, filtering, transformation, testing, modeling, visualization and documentation, and the result is outputted (or) the data is stored accordingly in data warehouse or databases.

Smart farming technologies can include the following methods for data collection, data process and decision making, and application technologies.

- Data collection technology involves:
    - Soil sampling and applications
    - Yield monitoring and mapping
    - Global satellite positioning
    - Remote sensing
    - Field and crop scouting
- Data process and decision-making technologies covers:
    - Geographical systems
    - Agricultural mapping software
    - Economic analysis
    - Geo-statistics
    - Modeling

- Application technologies focus on:
    - Variable rate application
    - Section control in farms
    - Agricultural drones & robots

The description of each unit in this framework as follows:

- **Fetch the Data and Load the Data to Transform Onto the Warehouse System:** The farming related data derived from various sources, such as sensors in the field, information from producers, as well as from aerial means such as satellite images, aerial photos and other remote sensing data.
- **Store and Use the Data in the Database System:** Such data need will be collected, processed, combined and exploited in a way that allows farmers to get exactly what they need: customized advice that will guide their cultivation practices based on data filtered through the scientific knowledge of researchers and the practical experience of agronomists / production advisors.
- Make available to access data for researchers for analytics.
- Examine the required data using suitable software.
- **Formulate the Data Visualization to Represent Data in Useful Format:** The farmer can have a precise overview of agriculture practice that should be followed.

Figure 7. Smart farming framework with data mining techniques

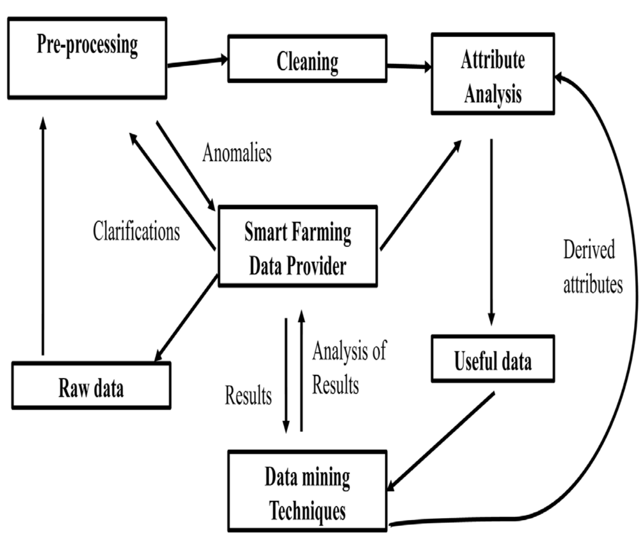
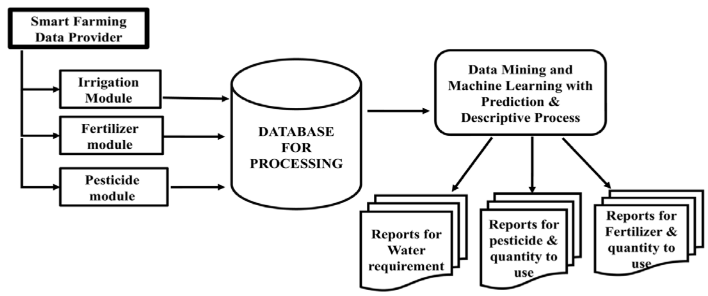


Figure 8. Data mining techniques conceptual framework for smart farming applications



Conceptual Framework

After considering the research that's been going on in the field of data mining approaches to extract information from the smart farming data provider, the system has been proposed to include machine learning concepts with data mining to apply for smart farming. The information of interest in the database includes details like crop name, crop stages, soil type, soil chemical condition, evaporation rate, temperature, humidity, growing period and disease symptoms. The conceptual framework (Figure 8) will provide an interface for the farmer to upload records of smart data providers. On this database, data mining and machine learning techniques can be applied to extract hidden patterns. Also, this system will be able to generate different kinds of reports as per the needs of the different farmers and researchers.

The proposed framework involves the three main modules:

1. **Fertilizer Module:** This module will be used to identify the fertilizer name and amount of fertilizer to be used on a particular crop and a particular stage of development. The data set consists of crop name, crop stage, soil chemical condition and growing period. History of the crop will be kept in the data set. The soil testing information is mandatory to identify the proper amount of fertilizer to use according to available nutrient, crop stage, etc.
2. **Irrigation Module:** In irrigation module water required per hector is identified. Irrigation depends on certain factors such as Crop name, Crop stage, soil condition, current day and next day temperature, Evaporation rate, humidity, etc. All these inputs are processed then give the result as water required per hector. Water is a limited resource and its conservation is the biggest crisis nowadays but using this system will aid into proper utilization of water & no wastage or under-over supply.
3. **Pesticide Module:** It is the key module for quality control & prevention of diseases. It is necessary to apply the proper amount of pesticide otherwise the excess or incorrect use of pesticide causes side effect on the crop. The factors like humidity, temperature, a sudden change in environment, excess use of water, water pollution, air pollution, soil pollution due to industries around the farm, etc. are considered for deciding the pesticide to be used for the crop. After processing, the relevant pesticide will be suggested to the farmer with their amount and per day use.

Because of autonomous in nature, this system won't need any possible interventions or time-to-time handling or changing of the data. It will precisely focus on the growth & cultivating of crops and will increase productivity by applying its tactics. This data will be uploaded to a central server to develop an integrated database. This database will be growing daily with new records being added to it. Using machine learning & data mining algorithms, models will be developed to extract patterns of interest and subsequently help in agricultural data prediction and corrective strategies using data analytics. Summarizing all the smart farming system is an idol solution for future farming. The authors are working to realize the models as discussed above.

APPLICATIONS OF DATA MINING TECHNIQUES IN SMART FARMING

Smart farming technologies require more and more professional skills to analyze agricultural data; extracting useful information has become the question of great importance. Smart agriculture is abundant with diverse information, which conditions the necessity to use data mining. The basic idea is when the same data analyzed in a different context, there will be a generation of new context-based information and knowledge. (Tejas S, 2015) suggested that the researchers design and develop applications for solving complex agricultural problems using data mining like yield data prediction, soil mapping, and classification, fertilizer and pesticide management, grading and sorting of agriculture product, etc.

There are several applications of data mining techniques in the field of agriculture and this section presents the most used data mining techniques. The data mining techniques used for automated prediction and analysis of various trends and behaviors, automated discovery of historical patterns dynamically, disease detection, pattern recognition, etc. by using multiple applications. Data mining techniques can get the benefits of automation on existing software and hardware platforms implemented on new systems with upgradation and the development of new products.

Forecasting Weather Conditions

- Weather forecasting using data mining on cloud computing with artificial neural network and decision tree algorithms (Wang, & Mujib 2017).
- Using the Hidden Markov Model for prediction and for extraction of the weather condition observations the K-means clustering (Rohit Kumar Yadav and Ravi Khatri, 2016).
- Classification and prediction of future weather forecasting by using the back propagation method (D.Sanjay et al. 2012)
- Weather prediction using the machine-learning technique with support vector machines (M.Shashi, 2009).
- Forecasting of water consumption in agriculture (Lu et al., 2009).
- Rainfall prediction using ANN and ARIMA techniques (V.K. Somvanshi, et al, 2006)
- Analyzing the changes in the weather scenarios (Tripathi et al. 2006) using SVMs.
- Independent component analysis to identify patterns of weather data (Basak et al., 2004).
- Forecasting the pollution in the atmosphere (Jorquera et al. 2001) using the k-means method.
- Simulating daily precipitations and other weather variables (Rajagopalan and Lall, 1999) using k-nearest neighbor.

Analysis of Soil and Water Characteristics

- Investigating soil supplements using classification algorithms (E.Manjula et al. 2017).
- Soil fertility prediction using classification algorithms such as Random forest, J48, Naïve Bayes (P. Jasmine Sheela, K. Sivaranjani, 2015)
- Classification of a soil sample to predict fertility rate using J48, Naïve Bayes, and Random forest (Vrushali Bhuyar, 2014).
- Soil classification using Naïve Bayes classifier (V. Ramesh and K. Ramar, 2011).
- Predicting soil fertility using a decision tree algorithm (Jay Gholap, 2012).
- Classify soils that analyze large soil profile experimental datasets using Naive Bayes (Bhargavi, P, 2009).
- Estimating soil water parameters and climate forecasting (Mucherino, A., Papajorgji, P., & Pardalos, P., 2009)
- Fuzzy set and interpolation techniques applied for land suitability evaluation for maize in Northern Ghana (Braimoh, A.K., Paul L. G. Vlek and Alfred Stein, 2004).
- Unsupervised and supervised methods for characterizing soil and canopy density in agricultural fields (Bajwa et al., 2004).
- Classifying soils and plants (Meyer et al. 2004) using a k-means approach
- Classifying soils in combination with GPS-based technologies (Verheyen et al. 2001) using k-means.
- Forecasting of water resources variables (Maier, H. R., & Dandy, G. C. 2000).
- Simulation models of soil dynamics (Jones et. al 1998).
- Classifying soils in combination with GPS-based technologies K-means Approaches using (Stockle et al., 1994).

Vegetable, Crop, and Fruit Cultivations

- Farmers' perception of risks in fruits and vegetable production (J, Ali & Kapoor, Sanjeev. (2019).
- Unay et al. (2011) used the automatic data mining techniques for recognizing and grading fruits.
- Yield estimation and clustering of chickpea genotypes using soft computing techniques (J.Khazaei, 2008).
- Rice yield prediction in mountainous regions using artificial neural network (Ji & Wan, 2007).
- Grading fruits by analyzing color images of fruits before marketing (Leemans & Destain 2004) using k-means approach.
- Monitoring the presence of water cores in X-ray images of fruits using a neural network for discriminating between good and bad apples (Shahin et al. 2001).
- Prediction of flowering and maturity dates of soybean (Elizondo, et al. 1994).

Crop Yield Prediction

- Using various data mining techniques used for predicting the crop yield (Y. Gandge & Sandhya, 2017).
- Plant nutrient management system to meet the needs of the soil, maintain its fertility levels, and hence improve the crop yield using machine learning methods (Shivnath Ghosh, 2014).

- Agricultural crop yield prediction using the ANN model (Snehal, S. D., 2014).
- (D Ramesh, 2013) applied k-means approach to estimate the crop yield analysis.
- Crop prediction framework for climate related variables for planting practices, particularly by the application right amount of fertilization (Rossana MC, 2013).
- Crop selection based on the effect of natural calamities like famines based on machine learning (Washington Okori, 2011).
- The use of artificial neural networks to choose crops based on soil and climate has been shown by researchers (Obua, 2011).
- Decker's high-resolution continuous soil classification using morphological soil profile descriptions (Nidhi Dwivedy, 2011)
- Support Vector Machine technique for classification of crops (Babu et al., 2010) and yield prediction (Ruß, 2009).
- Crop selection method which helps in crop selection based on its yield prediction and other factors (Kumar, 2009).
- Generalized Regression Neural Networks (GRNN) method is used for forecasting of agricultural crop production (Chaochong, 2008).
- Classifying crops (Camps-Valls et al. 2003) using SVMs.
- (Lee, & Kerschberg, 1998) used the knowledge discovery life cycle (KDLC) model for the study dealing with crop yield and visualization using GPS.

Plant Disease, Pesticide, and Fertilizers Control

- Controlled addition of fertilizers in order to avoid excess and deficient fertilizers in the soil (Amrutha et al., 2016).
- Artificial Neural Network approach for precise forecasting and forewarning models of plant diseases (Bhagawati et al., 2015).
- Recommendation of fertilizers based on soil fertility using compact decision tree (Navneet & Nasib Singh Gill, 2014).
- Using Regression analysis and cluster analysis to estimate the influence of fertilizer nutrients consumption on the wheat crop yield (Vijaykumar & Rajinikanth, 2013).
- Recommendation of the fertilizer (Baskar et al. 2013) using classification algorithms such as regression, J48, Naive Bayes.
- (Ding et al., 2010) proposed an application in the prediction of foodborne disease outbreaks (Thakura et al. 2010).
- Hierarchical clustering and association rule mining used to analyze pest scouting, pesticide and climatological parameters for optimization of pesticide usage and better management. (Tripathy et al., 2009; Abdullah et al., 2004).

Miscellaneous Applications

- Analysis of agriculture data using data mining techniques (Majumdar, 2017).
- Data mining techniques like Neural Networks, Fuzzy Logic, Evolutionary Algorithms, Bayesian Network and Support Vector Machine for grading agriculture products like apple, mango, strawberry, cherries, orange and etc. (Gill, Sandhu, & Singh 2014).

- Partitioning Algorithms and Hierarchical Algorithms) author examines the current usage and details of agriculture land vanished in the past seven years (Megala, 2011).
- Prediction of problematic wine fermentations (Mucherino, 2010) using supervised biclustering techniques.
- Detecting weed and nitrogen stress in corn using SVM (Karimi et al.2006).
- Monitoring water quality changes (K.A. Klise, 2006).
- Machine-learning techniques (Jain et al., 2009), Fuzzy Logic (Andujar, & Aroba, 2006; Meyer et al., 2004) for remote sensing, mapping, crop management, weed, and pest control for PA.
- Monitoring of wine fermentation process to be classified using ANNs (Riul et al. 2004).
- Holmgren and Thuresson (1998) used k-nearest neighbor approach to evaluate forest inventories and to estimate forest variables analyzing satellite imagery.

ISSUES OF DATA MINING APPLICATIONS IN SMART FARMING

The popularity and growth of data mining have also led to an increased concern about sustainable agriculture with the help of SF. Data mining requires a certain technology and analytic techniques, as well as the systems for reporting and monitoring which can measure results. The issues in the scope and complexity of agricultural data are one of the obstacles for successful data mining. Furthermore, information recorded in different formats from different data sources creates a large obstacle for successful data mining as it has missing, invalid, inconsistent or nonstandard data. It is very difficult for people to process gigabytes of records, although work with images is relatively easier because they are capable of recognizing patterns, accept basic trends in data, and formulate rational decisions. The stored information is becoming less useful if they are not in an easily understandable format.

Agricultural data often connected with uncertainty because of the inaccuracy of assessment, sample inconsistency, outdated data. Over the last few years (Sharma and Mehta, 2012), numerous studies have been conducted on managing uncertain data in databases, like presenting uncertainty in a database and finding data which contain uncertainty. However, few studies conducted regarding problems of researching uncertain data. To apply traditional techniques of data mining, uncertain data reduced to atomic values. Inconsistencies in acquired and real values can significantly influence the quality of results in research (Sharma and Mehta, 2012).

In addition, the major problem in the domain of data mining with agriculture can be limited in access to data, since raw input for data mining often exists in different settings and systems like agricultural professional services, services of the ministry of agriculture, laboratories, and other. Therefore, acquired data should be integrated before applying data mining techniques. Construction of data storage for data mining can be a very expensive and time-consuming process. Developing data mining applications in agricultural organizations involves huge investment resources, especially time, effort and money.

Demographic trends, including aging populations and continued migration of people from rural to urban areas, have attracted the attention of researchers, because labor issues may become a scarcity factor in agriculture. In addition to these trends, the intensification of climate change will continue to alter growing conditions, such as the temperature, precipitation, and soil moisture, in less predictable ways (Brazilian Agricultural Research Corporation, 2014). These impacts can be reduced by SF tools and they can assist in minimizing environmental constraints (S. Fountas, 2015).

Awareness and adoption rate of PA technologies are affected by many factors including characteristics of the farms, personality and family structure of the farmer, features of the equipment, characteristics of the technology, legal affairs, social interaction, etc. Farmers adopt and use PA technologies for specific benefits (Say & Sait, 2017).

(DEFRA, 2013) surveyed farmers in England and reported they use PA technologies mostly for improving accuracy, reducing input costs, improving soil conditions, improving operator conditions and reducing greenhouse gas emissions. Also, reasons for not using PA included being not cost effective and/or high initial setup costs, being not suitable for type or size of the farm, being too complicated to use, and not accurate enough (DEFRA, 2013).

Keskin et al. (2016) reported that the farmers using tractor auto guidance in the Adana province of Turkey used PA technology as creating straight crop rows, flexible working hours, time-saving, fuel saving, labor saving agricultural input saving and yield increase.

Development of ICT in agriculture has enabled the creation of electronic record about farmers that acquired by monitoring customers. These electronic records include a series of information such as farmer's demographics, sales progress records, crop examination details, used protective equipment, previous crop rotation data, pedological findings. Information system automates and simplifies the workflow of agricultural enterprises.

Integration of data mining in agricultural organization's information systems reduces subjectivity in decision making and provides new useful agricultural knowledge using predictive and descriptive models provide the best support of knowledge and experience to agricultural workers. Important questions arise here who need answering; what the relationship between attributes and outcomes is; what the interesting relationship is be found between attributes; which data are corresponding to predictive characteristics enough to build a predictive model of acceptable performance; which immediate factors be extracted from original attributes which can increase performance of the predictive model.

Agricultural organizations developing data mining requires high investment resources, especially time, effort and money. A data mining project can fail for a number of reasons, like lack of support from management, inadequate data mining expertise, the disinterest of agronomists and professional services.

SOLUTIONS AND RECOMMENDATIONS

Farming is highly unpredictable, due to its large dependency on weather and environmental conditions such as temperature, rain, humidity, hail, unpredictable events like plant diseases, pests, and price volatility in agricultural markets. This implies the need for sensor and automation technologies, data analytics with large-scale agriculture-related frameworks, in order to help the farmers, become informed about their farms' conditions and risks early enough, in order to take proper counter-measures and protect their crops and overall production. Data mining acknowledged as the most advanced concept for prediction of market fluctuation and price variability.

With IoT fascinated SF, there is no room for the guesswork in farming because soil sensors used across the farmland can alert the farmer to any irregular conditions like high acidity or low moisture. The farmers can get accurate soil data either by the dashboard or a customized mobile application.

In agriculture, physical parameters such as temperature, relative humidity, and soil moisture are important (Zhang, N., Wang, M., & Wang, N. (2002). Several applications and well-established measuring instruments are used to collect these data (Schellberg, J., et al., 2008). In addition, sensors to measure soil properties (Cox, 2002), fertilizer management (Corwin & Lesch, 2005) detect and monitor foliar disease (Adamchuk, V. I., Hummel, J. W., Morgan, M. T., & Upadhyaya, S. K., 2004) already exist.

McQueen, et al. (1995) and Gebbers and Adamchuk (2010) suggested that to address the challenges of SF and sustainable agriculture, there is a necessity for better analysis and understanding of the complex, multivariate and unpredictable agricultural ecosystems. The emerging digital technologies contribute to this understanding by monitoring and measuring continuously various aspects of the physical environment (Sonka, 2016), producing large quantities of data in an unprecedented pace (Chi, et al., 2016). This implies as pointed out by (Hashem, et al., 2015) there is a need for large-scale collection, storage, preprocessing, modeling and analysis of huge amounts of data coming from various heterogeneous sources.

Business Intelligence (SF Challenges) predicts that the various agriculture IoT devices installations will hit 75 million by 2020, growing 20% annually and the global smart agriculture market size expected to $15.3 billion by 2025.

Telecom operators can support IoT deployment in agriculture at multiple levels by providing good connectivity provisioning, authentication, security, billing, device management, location-based services, application enablement, and analytic services.

To maximize the crop yield, selection of the appropriate crop that sown plays a vital role. It depends on various factors like the type of soil and its composition, climate, the geography of the region, crop yield, market prices, etc. Techniques like ANN, K-nearest neighbors and Decision Trees have carved a niche for themselves in the context of crop selection which is based on various factors. The applications that use the k-means approach, utilize only the basic algorithm, while many other improvements are available in the literature (Hansen and Mladenovic 2002; Jinlan et al. 2005). The reason for using only the basic version of the techniques may be related to the availability of only the basic expertise in computer science and numerical analysis of researchers in agriculture-related fields. It is better to handle the complex projects with a multi-disciplinary team, where mathematicians and computer scientists are part of the team.

An accurate forecast of weather can reduce the enormous toil faced by farmers in India including crop selection, watering, and harvesting. As the farmers have poor access to the Internet as a result of digital-divide, they have to rely on the little information available regarding weather reports. Up-to-date as well as accurate weather information is still not available as the weather changes dynamically over time. Researchers have been working on improving the accuracy of weather predictions by using a variety of algorithms like ANN adopted extensively for this purpose.

Data mining techniques can solve the problem of predicting yield production by collecting the sensor data which are available for some time back to the past, where the corresponding yield productions have been recorded. All this information forms a training set of data which can be exploited to learn how to classify future yield productions, once new sensor data are available.

FUTURE TRENDS AND RESEARCH DIRECTIONS

Smart farming is vital for the future of agriculture. Knowledge and capital are essential for any innovation. More skill and professionalism are required for new farming technologies. A farmer today is not only a person with a passion for agriculture, but he should be a legal expert to find their way through a growing maze of regulations and a part-time data analyst, economist and accountant with bookkeeping skills for making and selling agricultural produce an in-depth knowledge of market chains and price volatility. Furthermore, SF needs capital, for instance, smartphone applications that help sustainable agriculture are available in the market, which shows that implementing SF technologies easily upscale.

While the data mining applications provide new features and benefits to economic and efficient use of resources, it requires timely and sophisticated analysis of an integrated view of the data. In the era of hi-tech agriculture, there is a host of opportunities for technologists to completely transform the farming ecosystem. These smart technologies provide extremely productive ways to improve the soil quality, increase crop production with the use of affordable sensors and then do well-informed decisions based on data analytics, which in turn would improve operational efficiency. Thus, with the installation of SF technologies, both small farmers and large landowners can increase efficiency and sustainability in their crop productions. The yield of agriculture primarily depends on diseases, pests, climatic conditions, planning of different crops for the harvesting productivity are the results. Therefore, these predictions are very useful for agriculture domains.

PA research has the ability to define a decision support system with the goal of optimizing returns on inputs while preserving resources. (Ess & Morgan, 2003; Rains & Thomas, 2009) cited the technological developments in the agricultural sector yield better management practices resulting in more precision in the agricultural operations from tillage to harvesting to reduce inputs, increase profits, and protect the environment.

The technology of data mining can generate new business opportunities by providing these capabilities and the researchers can focus on the following areas:

- Automated discovery of historical patterns dynamically.
- Automation on existing software and hardware platforms to support new systems.
- Developments of high-performance parallel systems.
- System for explaining pesticide abuse by data mining.
- Making a novel pilot agriculture extension data warehouse.
- Automated analysis and prediction of various trends and behaviors.
- Techniques for capture, prepare and store data meticulously.
- The supportive platform for creating a community of practitioners.

Farmers and agriculture companies are turning to the IoT to satisfy the need for the population and encouraging the future of farming to the next level. Smart agriculture is becoming more commonplace among farmers, and high-tech farming is quickly becoming the standard because of the agricultural sensors. Collecting big data from the farm field gives a richer understanding of product variability and quality of products (Camilli, A., Cugnasca et. al., 2007).

In a near future, more techniques that are sophisticated tailored to address complex problems in agriculture-related fields and hence provide better results which evidence in the growing number of DM techniques in agriculture applications and a growing amount of data that are currently available from many resources. Moreover, researchers can often cross-disciplinary lines with their work by combining biology, economics, engineering, chemistry, community development, and many others.

Systems that are more powerful developed with advanced programming methodologies to design and develop complex agricultural systems (Papajorgji & Pardalos 2006) which require advanced skills in computer science. Some data mining techniques like biclustering techniques have not yet been applied to agricultural problems, which may be useful for discovering important information from agricultural-related sets of data.

There is a huge database of agriculture is available for data mining techniques. There are a lot of opportunities for research on data mining in agriculture to develop practical applications for the enhancing of agriculture production. If more complex techniques implemented with high-performing computational systems for data mining will surely solve difficult problems in the agriculture sector. There are some researches on agriculture like the application of spatial data mining, developing an innovative application in agriculture, classification of land soils and a tool for a knowledge management system in agriculture. The researcher can also analyze many factors in agriculture using some data mining techniques to increase crop yields thereby increasing the farmer's economy and status of our country in the ancient period.

CONCLUSION

Due to the advent of smart farming techniques, farming has seen a number of technological transformations in the last decades, becoming more industrialized and technology-driven. By using various smart agriculture gadgets, farmers have gained better control over the process of growing crops, making it more predictable and improving its efficiency. Smart farming involves the use of modern technologies in the farm management cycle. In addition, data mining techniques draw work from areas including database technology, machine learning, statistics, pattern recognition, information retrieval, neural networks, knowledge-based systems, artificial intelligence, and high-performance computing and data visualization. Compare to the conventional data analysis techniques used in agricultural research, Data mining can open a new avenue for research and development in agriculture and associated ventures, also focusing on issues relating to their practicality, usefulness, feasibility, and scalability. The proposed system works on the data mining and machine learning techniques along with the data obtained from satellite information, Internet, from soil testing report fed in the existing databases. This system covers irrigation system to make the decision about the water requirement for the crop at particular day or stage, fertilizer system provides the recommendation of fertilizer to be used at particular stage according to the nutrients in the soil, pesticide system provides the quantity of pesticide to be used depend on various environmental factors. This system is capable of increasing the productivity of fields by managing farm operations smartly. The multidisciplinary of the research teams is very important because mathematicians and computer scientists can help agronomists in finding complex solutions to these complex problems. It is believed that the use of more complex techniques, such as biclustering, and high-performing computational systems, such as parallel computers, in data mining, will help to solve complex problems in agriculture-related fields for sustainable agriculture. In conclusion, there is a lot of work to be done on this emerging and interesting research field in the near future.

REFERENCES

Abdullah, A., Brobst, S., Pervaiz, I., Umer, M., & Nisar, A. (2004). Learning dynamics of pesticide abuse through data mining. *Proc. of Australian Workshop on Data Mining and Web Intelligence*.

Abello, J., Pardalos, P. M., & Resende, M. (2002). *Handbook of massive data sets*. New York: Kluwer. doi:10.1007/978-1-4615-0005-6

Adamchuk, V. I., Hummel, J. W., Morgan, M. T., & Upadhyaya, S. K. (2004). On-the-go soil sensors for precision agriculture. *Computers and Electronics in Agriculture*, *44*(1), 71–91. doi:10.1016/j.compag.2004.03.002

Ahmed, S. R. (2004). Applications of data mining in retail business. *International Conference on Information Technology: Coding Computing, 2*, 455-459. 10.1109/ITCC.2004.1286695

Ali & Kapoor. (2019). Farmers' perception on risks in fruits and vegetables production: an empirical study of Uttar Pradesh. *Agricultural Economics Research Review*, 21.

Amrutha, A., Lekha, R., & Sreedevi, A. (2016). *Automatic Soil Nutrient Detection and Fertilizer Dispensary System*. IEEE. doi:10.1109/RCTFC.2016.7893418

Andujar, M., Aroba, J., de Torre, M. L., & Grande, J. A. (2006). Contrast of evolution models for agricultural contaminants in ground waters by means of fuzzy logic and data mining. *Environmental Geology*, *49*(3), 458–466. doi:10.100700254-005-0103-2

Babinet, G. (2015). *The New World economy, s.l.: Report addressed to Ms Segolene Royal, Minister of Environment, Sustainable Development and Energy.* Working group led by Corinne Lepage.

Babu, Ramana Murthy, & Narayana. (2010). A web based tomato crop expert information system based on artificial intelligence and machine learning algorithms. *Int. J. of Comp. Sci., and Information Technologies*, *1*(1), 6-15.

Bajwa, S., Bajcsy, P., Groves, P., & Tian, L. F. (2004). Hyperspectral image data mining for band selection in agricultural applications. *Transactions of the ASAE. American Society of Agricultural Engineers*, *47*(3), 895–907. doi:10.13031/2013.16087

Bakhshinategh, B., Zaiane, O. R., ElAtia, S., & Ipperciel, D. (2018). Educational data mining applications and tasks: A survey of the last 10 years. *Education and Information Technologies*, *2*(1), 537–553. doi:10.100710639-017-9616-z

Basak, J., Sudharshan, A., Trivedi, D., & Santhanam, M. S. (2004). Weather Data Mining Using water/nitrogen budgets and crop yield. *Agricultural Systems*, *46*(3), 335–359.

Baskar, S. S., Arockiam, L., & Charles, S. (2013). *Applying Data Mining Techniques on Soil Fertility Prediction*. *IJCAT*. doi:10.7753/IJCATR0206.1005

Bastiaanssen, W., Molden, D., & Makin, I. (2000). Remote sensing for irrigated agriculture: Examples from research and possible applications. *Agricultural Water Management*, *46*(2), 137–155. doi:10.1016/S0378-3774(00)00080-9

Beniwal, S., & Arora, J. (2012). Classification and feature selection techniques in data mining. *International Journal of Engineering Research & Technology*, *1*(6).

Berry, M. J. A., & Linoff, G. S. (2004). *Data mining technique for marketing, sales, and customer relationship management* (2nd ed.). Wiley.

Bhagawati, R., Bhagawati, K., Singh, A. K. K., Nongthombam, R., Sarmah, R., & Bhagawati, G. (2015). Artificial Neural Network Assisted Weather Based Plant Disease Forecasting System. *International Journal on Recent and Innovation Trends in Computing and Communication*, *3*(6), 4168–4173.

Bhargavi, P, & Jyothi, S. (2009). Applying Naive Bayes data mining technique for classification of agricultural land soils. *International Journal of Computer Science and Network Security*, *9*(8), 117-122.

Bhasin, M. L. (2006). Data mining: A competitive tool in the banking and retail industries. ML Bhasin. *Chartered Accountant-New Delhi*, *55*(4), 588.

Bhuyar, V. (2014). *Comparative Analysis of Classification Techniques on Soil Data to Predict Fertility Rate for Aurangabad District*. *IJETTCS*.

Bongiovanni, R., & Lowenberg-DeBoer, J. (2004). Precision agriculture and sustainability. *Precision Agriculture*, *5*(4), 359–387. doi:10.1023/B:PRAG.0000040806.39604.aa

Braimoh, Vlek, & Stein. (2004). Land Evaluation for Maize Based on Fuzzy Set and Interpolation. *Environmental Management*, *33*(2), 236–238. PMID:15285400

Braimoh, A. K., & Vlek Paul, L. (2004). Land Evaluation for Maize Based on Fuzzy Set and Interpolation. *Environmental Management*, *33*(2), 226–238. doi:10.100700267-003-0171-6 PMID:15285400

Brajovic, Vujovic, & Dukanovic. (2015). An Overview of Smart Irrigation Software. *4th Mediterranean Conference on Embedded Computing, MECO*.

Brazilian Agricultural Research Corporation. Visão 2014–2034. (2014). Retrieved May 17, 2019, from Link: https://www.embrapa.br/documents/1024963/1658076/Documento+Visão+-+versão+completa/7bf520f2-7329-42c0-8bf0-15b3353c3fdb

Brettel, Friederichsen, Keller, & Rosenberg. (2014). How virtualization, decentralization and network building change the manufacturing landscape: an industry 4.0 perspective *IJMEA*, *8*(1), 37-44.

Camilli, A., Cugnasca, C. E., Saraiva, A. M., Hirakawa, A. R., & Corrêa, P. L. P. (2007). From wireless sensors to field mapping: Anatomy of an application for precision agriculture. *Computers and Electronics in Agriculture*, *58*(1), 25–36. doi:10.1016/j.compag.2007.01.019

Camps-Valls, G., Gomez-Chova, L., Calpe-Maravilla, J., Soria-Olivas, E., Martin-Guerrero, J. D., & Moreno, J. (2003). Support vector machines for crop classification using hyperspectral data. *Lecture Notes in Computer Science*, *2652*, 134–141. doi:10.1007/978-3-540-44871-6_16

Challa, M. (2013). *Determining Factors and Impacts of Modern Agricultural Technology Adoption in West Wollega, Munich*. GRIN Publishing GmbH. Retrieved May 15, 2019, from://www.grin.com/en/e-book/280336/determiningfactors-and-impacts-of-modern-agricultural-technology-adoption

Chandola, V., Banerjee, A., & Vipin, K. (2009). Anomaly detection: A survey. *ACM Computing Surveys*, *41*(3), 1–58. doi:10.1145/1541880.1541882

Chaochong, J. M. (2008). *Forecasting Agricultural Production via Generalized Regression Neural Network*. IEEE.

Chi, M., Plaza, A., Benediktsson, J. A., Sun, Z., Shen, J., & Zhu, Y. (2016). Big data for remote sensing: Challenges and opportunities. *Proceedings of the IEEE*, *104*(11), 2207–2219. doi:10.1109/JPROC.2016.2598228

Sawaitul, Wagh, & Chatur. (2012). Classification and Prediction of Future Weather by using Back Propagation Algorithm-An Approach. *International Journal of Emerging Technology and Advanced Engineering*, *2*(1).

Corwin, D. L., & Lesch, S. M. (2005). Apparent soil electrical conductivity measurements in agriculture. *Computers and Electronics in Agriculture*.

Cox, S. (2002). Information technology: The global key to precision agriculture and sustainability. *Computers and Electronics in Agriculture*, *36*(2-3), 93–111. doi:10.1016/S0168-1699(02)00095-9

Cunningham & Holmes. (1999). *Developing innovative applications in agriculture using data mining*. Academic Press.

Dass. (n.d.). *Data Mining in Banking and Finance: A Note for Bankers*. Indian Institute of Management Ahmadabad.

DEFRA. (2013). Farm Practices Survey Autumn 2012 – England. Department for Environment, Food and Rural Affairs (DEFRA), 41.

Desai & Desai. (2004). The Role of Data mining in Banking Sector. *IBA Bulletin*.

Ding, L., Meng, J., & Yang, Z. (2010). An Early Warning System of Pork Price in China Based on Decision Tree. *IEEE Conference Proceedings, International Conference on E-Product E-Service and EEntertainment*, 1–6. 10.1109/ICEEE.2010.5660808

Dubey, Kamath, & Kanakia. (2016). Learning Data Mining Techniques. *International Journal of Computer Applications*.

Dwivedy. (2011). Challenges faced by agriculture sector on developing countries with special reference India. *International Journal of Rural Studies, 18*(2).

Eapen. (2004). *Application of Data mining in Medical Applications*. UWSpace. Retrieved from http://hdl.handle.net/10012/772

Elizondo, D. A., McClendon, R. W., & Hoogenboom, G. (1994). Neural network models for predicting flowering and physiological maturity of soybean. *Transactions of the ASAE. American Society of Agricultural Engineers*.

Emre, I. E., Erol, N., Ayhan, Y. İ., Özkan, Y., & Erol, Ç. (2019). The analysis of the effects of acute rheumatic fever in childhood on cardiac disease with data mining. *International Journal of Medical Informatics*, *123*, 68–75. doi:10.1016/j.ijmedinf.2018.12.009 PMID:30654905

Everitt, B., Landau, S., & Leese, M. (2001). *Cluster Analysis*. London: Arnold.

Fayyad, Piatetsky-Shapiro, & Smyth. (1996). *Advances in knowledge discovery and data mining*. Academic Press.

Fountas, S., Pedersen, S.M., & Blackmore, S. (2005). ICT in Precision Agriculture – diffusion of technology. ICT in agriculture: Perspectives of Technological Innovation, 15.

Fountas, S., Carli, G., Sørensen, C. G., Tsiropoulos, Z., Cavalaris, C., & Vatsanidoud, A. (2015). Farm management information systems: Current situation and future perspectives. *Computers and Electronics in Agriculture*, *115*, 40–50. doi:10.1016/j.compag.2015.05.011

Gandge, Y. & Sandhya. (2017). A study on various data mining techniques for crop yield prediction. *International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT)*, 420-423. 10.1109/ICEECCOT.2017.8284541

Gebbers, R., & Adamchuk, V. (2010). Precision agriculture and food security. *Science*, *327*(5967), 828–831. doi:10.1126cience.1183899 PMID:20150492

Gholap, J. (2012). Performance tuning of j48 algorithm for prediction of soil fertility. *Asian Journal of Computer Science and Information Technology*, *2*(8), 251–252.

Gill, J., Sandhu, D. S., & Singh, D. (2014). A Review of Automatic Fruit Classification using Soft Computing Techniques. In *International Conference on Computer, Systems and Electronics Engineering (ICSCEE'2014),* (pp. 91-98). Johannesburg, South Africa: Grand View Research. Retrieved May 17, 2019, from https://www.grandviewresearch.com/press-release/global-agriculture-drones-market

Grossman, R., Kamath, C., Kegelmeyer, P., Kumar, V., & Namburu, R. (2001). *Data Mining for Scientific and Engineering Applications*. doi:10.1007/978-1-4615-1733-7

Han, J., Kamber, M., & Pei, J. (2006). *Data mining: concepts and techniques*. Morgan kaufmann.

Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufman Publishes.

Hashem, I., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Ullah Khan, S. (2015). The rise of "big data" on cloud computing: Review and open research issues. *Information Systems*, *7*, 98–115. doi:10.1016/j.is.2014.07.006

Hinton, G., Osindero, S., & Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, *18*(7), 1527–1554. doi:10.1162/neco.2006.18.7.1527 PMID:16764513

Hussain, M., Gawate, S. P., Prasad, P. S., & Kamble, P. A. (2015). Smart irrigation system with three level access mechanisms. *International Conference on Computation of Power, Energy, Information and Communication (ICCPEIC)*. 10.1109/ICCPEIC.2015.7259474

ILEIA. (2000). Grassroots innovation. Unleashing the creativity of farmers. *Newsletter for Low External Input and Sustainable Agriculture*, *16*(2), 4–5.

Jain, A., & Dubes, R. (1998). *Algorithms for Clustering Data*. Englewood Cliffs, NJ: Prentice-Hall.

Jain, R., Minz, S., & Rama Subramaniam, V. (2009). Machine learning for forewarning crop diseases. *Journal of the Indian Society of Agricultural Statistics*, *63*(1), 97–107.

Jasmine Sheela & Sivaranjani. (2015). A Brief Survey of Classification Techniques Applied To Soil Fertility Prediction. *ICETSH-2015*.

Ji, B. S., & Wan, J. (2007). Artificial neutral networks for rice yield prediction in mountainous regions. *Journal of Agricultural Science*.

Jones, J. W. (1998). Decision support system for agrotechnology transfer: DSSAT v3. In G. Y. Tsuji, G. Hoogenboom, & P. K. Thornton (Eds.), *Understanding options for agricultural production* (pp. 157–177). Kluwer Academic Publishers. doi:10.1007/978-94-017-3624-4_8

Jorquera, H., Perez, R., Cipriano, A., & Acuna, G. (2001). Short term forecasting of air pollution episodes. In P. Zannetti (Ed.), *Environmental modeling 4*. WIT Press.

Kamesh, V. (2019). Real-Time Fraud Anomaly Detection in E-banking Using Data Mining Algorithm. *South Asian Journal of Engineering and Technology*, *8*(1), 144–148.

Kamilaris, A., Gao, F., Prenafeta-Boldú, F. X., & Ali, M. I. (2016). *Agri-IoT: A Semantic Framework for Internet of Things-enabled Smart Farming Applications*. Reston, VA: Academic Press.

Kaptan, S. S. (2002). *New Concepts in Banking*. Sarup & Sons, Edition.

Karmas, A., Karantzalos, K., & Athanasiou, S. (2014). Online analysis of remote sensing data for agricultural applications. s.l. *OSGeo's European conference on free and open source software for geospatial*.

Karugia, S., Baltenweck, I., Waithaka, M., Miano, M., Nyikal, R., & Romney, D. (2004). *Perception of Technology and its Impact on Technology Uptake: The Case of Fodder Legume in Central Kenya Highlands*. The Role of Social Scientists Proceedings of the Inaugural Symposium, Nairobi, Kenya.

Keskin, M., & Sekerli, Y. E. (2016). Awareness and adoption of precision agriculture in the Cukurova region of Turkey. *Agronomy Research (Tartu)*, *14*(4), 1307–1320.

Khazaei, J., Naghavi, M. R., Jahansouz, M. R., & Salimi-Khorshidi, G. (2008). Yield estimation and clustering of chickpea genotypes using soft computing techniques. *Agronomy Journal*, *100*(4), 1077. doi:10.2134/agronj2006.0244

Klise, K. A., & McKenna, S. A. (2006). *Water Quality Change Detection: Multivariate Algorithms. Proceedings of SPIE 6203*.

Klosgen, W., & Zytkow, J. M. (2002). *Handbook of data mining and knowledge discovery*. Oxford University Press.

Kumar, R. M. (2009). Crop Selection Method to Maximize Crop Yield Rate Using Machine Learning Technique. *Energy Materials*, *1*(1).

Lashari, Ibrahim, Senan, & Taujuddin. (2018). Application of Data Mining Techniques for Medical Data Classification: A Review. *MATEC Web of Conferences, 150*, 06003.

Lasi, H., Fettke, P., Kemper, H. G., Feld, T., & Hoffmann, M. (2014). Industry 4.0. *Business & Information Systems Engineering*, *6*(4), 239–242. doi:10.100712599-014-0334-4

Lee, S., & Kerschberg, L. (1998). Methodology and life cycle model for data mining and knowledge discovery in precision agriculture. *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, 3, 2882–2887.

Liao, Y., Deschamps, F., Loures, E. D. F. R., & Ramos, L. F. P. (2017). Past, present and future of industry 4.0-a systematic literature review and research agenda proposal. *International Journal of Production Research*, *55*(12), 3609–3629. doi:10.1080/00207543.2017.1308576

Loevinsohn, M., Sumberg, J., & Diagne, A. (2012). Under what circumstances and conditions does adoption of technology result in increased agricultural productivity? Protocol. London: EPPI Centre, Social Science Research Unit, Institute of Education, University of London.

Lokers, R., Knapen, R., Janssen, S., van Randen, Y., & Jansen, J. (2016). Analysis of Big Data technologies for use in agro-environmental science. *Environmental Modelling & Software*, *84*, 494–504. doi:10.1016/j.envsoft.2016.07.017

Lu, S., Cai, Z. J., & Zhang, X. B. (2009). Forecasting Agriculture Water Consumption based on PSO and SVM. In *IEEE International Conference on Computer Science and Information Technology* (pp. 147–150). Beijing, China: IEEE.

Maier, H. R., & Dandy, G. C. (2000). Neural networks for the prediction and forecasting of water resources variables: A review of modelling issues and applications. *Environmental Modelling & Software*, *15*(1), 101–124. doi:10.1016/S1364-8152(99)00007-9

Majumdar, J., & Naraseeyappa, S., & Ankalaki. (2017). Analysis of agriculture data using data mining techniques: Application of big data S. *J. Big Data*, *4*(20).

Manjula, E. (2017). Data mining technique to analyze soil nutrients based on hybrid classification. *International Journal of Advanced Research in Computer Science.*, *8*(8), 505–510. doi:10.26483/ijarcs.v8i8.4794

Manoj & Singhm. (2011). Applications of Data Mining for Intrusion Detection. *International Journal of Educational Planning and Administration*, *1*(1), 37–42.

Masoud, R., & Ahmed, T. M. (2016). Using Data Mining in Telecommunication Industry: Customer's Churn Prediction Model. *Journal of Theoretical and Applied Information Technology*, *91*(2).

Maynard, A. D. (2015). Navigating the fourth industrial revolution. *Nature Nanotechnology*, *10*(12), 1005–1006. doi:10.1038/nnano.2015.286 PMID:26632281

McQueen, R., & Garner, S., C.G., N.-M., & Witten, I. H. (1995). Applying machine learning to agricultural data. *Computers and Electronics in Agriculture*, *12*(1).

Megala, S., & Hemalatha, M. (2011). A Novel Datamining Approach to Determine the Vanished Agricultural Land in Tamilnadu. *International Journal of Computers and Applications*, 23. doi:10.5120/2869-3718

Meyer, G. E., Neto, J. C., Jones, D. D., & Hindman, T. W. (2004). Intensified fuzzy clusters for classifying plant, soil, and residue regions of interest from color images. *Computers and Electronics in Agriculture*, *42*(3), 161–180. doi:10.1016/j.compag.2003.08.002

Miller, McCarthy, & Zakzeski. (2009). A Fresh Approach to Agricultural Statistics: Data Mining and Remote Sensing. *Section on Government Statistics*, 3144-3155.

Mucherino, A., Papajorgji, P., & Pardalos, P. (2009). *Data mining in agriculture* (Vol. 34). Springer. doi:10.1007/978-0-387-88615-2

Mucherino, A., & Ruß, G. (2011). *Recent Developments in Data Mining and Agriculture*. Retrieved May 17, 2019, from http://www.antoniomucherino.it/download/papers/MucherinoRuss-dma11.pdf

Mucherino, A., & Urtubia, A. (2010). Consistent Biclustering and Applications to Agriculture. In *Proceedings of the Industrial Conference on Data Mining (ICDM10), Workshop "Data Mining in Agriculture" (DMA10)* (pp.105-113). Berlin, Germany: Academic Press.

Mwangi, M., & Kariuki, S. (2015). Factors Determining Adoption of New Agricultural Technology by Smallholder Farmers in Developing Countries. *Journal of Economics and Sustainable Development*, *6*(5), 208–216.

Namratha & Sharma. (2016). Educational Data Mining –Applications and Techniques. *International Journal of Latest Trends in Engineering and Technology*, *7*(2), 484–488.

Nathan, A. (2016). *Customer segregation in banking organization using knowledge management*. Academic Press.

Navneet & Gill. (2014). Algorithm for Producing Compact Decision Trees for Enhancing Classification Accuracy in Fertilizer Recommendation of Soil. *IJCA*.

Nguyen, H. A., & Choi, D. (2008). Application of Data Mining to Network Intrusion Detection: Classifier Selection Model. *APNOMS*. *LNCS Springer*, *5297*, 399–408.

Njenga, B. (2012). *The Application of Data Mining in the Mobile Telecommunications Industry in Kenya*. University of Nairobi.

Obua, W. O. (2011). Machine Learning Classification Technique for Famine Prediction. *Proceedings of the World Congress on Engineering*, 2.

Parack, S., Zahid, Z., & Merchant, F. (2012). Application of data mining in educational databases for predicting academic trends and patterns. *IEEE International Conference on Technology Enhanced Education (ICTEE)*, 1-4. 10.1109/ICTEE.2012.6208617

Pardalos, P. M., Boginski, L. V., & Vazacopoulos, A. (2007). *Data mining in biomedicine*. New York: Springer. doi:10.1007/978-0-387-69319-4

Pardalos, P. M., & Hansen, P. (2008). *Data mining and mathematical programming*. American Mathematical Society. doi:10.1090/crmp/045

Patel, Patel, & Upadhyay. (2016). A Survey Paper of Study about Data Mining. *International Journal for Research in Computer Science*.

Patil, Naik, Beldar, & Deshpande. (2016). Smart Farming using Ardino and Data Mining. *International Conference on Computing for Sustainable Global Development*.

Pivoto, D., Waquil, P. D., Talamini, E., Finocchio, C. P. S., Dalla Corte, V. F., & de Vargas Mores, G. (2018). Scientific development of smart farming technologies and their application in Brazil. *Information Processing in Agriculture*, *5*(1), 21–32. doi:10.1016/j.inpa.2017.12.002

Poultney, C., Chopra, S., & Cun, Y. L. (2006). Efficient learning of sparse representations with an energy-based model. *Advances in Neural Information Processing Systems*, 1137–1144.

Precision Agriculture. (2015). *What-is-precision-agriculture-and-why-is-it-important*. Retrieved May 15, 2019, from https://soilsmatter.wordpress.com/2015/02/27/what-is-precision-agriculture-and-why-is-it-important/2015

Pretty, J. (2008). Agricultural sustainability: Concepts, principles and evidence. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *363*(1491), 447–465. doi:10.1098/rstb.2007.2163 PMID:17652074

Rahman, H., Sudheer Pamidimarri, D., Valarmathi, R., & Raveendran, M. (2013). *Omics: Applications in Biomedical, Agriculture and Environmental Sciences, s.l.* CRC Press.

Rajagopalan, B., & Lall, U. (1999). A k-nearest-neighbor simulator for daily precipitation and other weather variables. *Water Resources Research*, *35*(10), 3089–3101. doi:10.1029/1999WR900028

Ramesh, D., & Vishnu Vardhan, B. (2013). Data Mining Techniques and Applications to Agricultural Yield Data. *International Journal of Advanced Research in Computer and Communication Engineering*, *2*(9).

Ramesh, V., & Ramar, K. (2011). Classification of Agricultural Land Soils: A Data Mining Approach. *Agricultural Journal*, *6*(3), 82–86. doi:10.3923/aj.2011.82.86

Reijntjes, C., & Waters-Bayer, A. (Eds.). (2001). *Farmer Innovation in Africa. A source of Inspiration for Agricultural Development*. London, UK: Earthscan Publications.

Ren, Z., & Lu, X. (2015). Design of fertilization recommendation knowledge base and application. *First International Conference on Agro- Geoinformatics (Agro-Geoinformatics)*, 1-5.

RIICE Partnership. (2014). *Remote sensing-based Information and Insurance for Crops in Emerging economies*. Retrieved May 15, 2019, from http://www.riice.org/

Rossana, M. C. (2013). A Prediction Model Framework for Crop Yield Prediction. *Asia Pacific Industrial Engineering and Management Society Conference Proceedings*, 185.

Ruß, G. (2009). Data Mining of Agricultural Yield Data: A Comparison of Regression Models. In Proceedings of Advances in Data Mining – Applications and Theoretical Aspects (pp. 24–37). Berlin: Springer.

Ruß, G., & Brenning, A. (2010). Data Mining in Precision Agriculture: Management of Spatial Information. Computational Intelligence for Knowledge-Based Systems Design. In *13th International Conference on Information Processing and Management of Uncertainty* (pp. 350-359). Dortmund, Germany: Academic Press.

Ryu, M. (2015). Design and implementation of a connected farm for smart farming system. *ICSENS*, *1-4*. doi:10.1109/ICSENS.2015.7370624

Saad, N. (2002). *Farmer processes of experimentation and innovation – A review of literature*. Cali, Colombia: Participatory Research and Gender Analysis Program, PRGA Working Document No. 21.

Say, S., Keskin, M., Sehri, M., & Sekerli, Y. (2017). *Adoption of Precision Agriculture Technologies in Developed and Developing Countries*. *International Science and Technology Conference*, Berlin, Germany.

Schellberg, J., Hill, M. J., Gerhards, R., Rothmund, M., & Braun, M. (2008). Precision agriculture on grassland: Applications, perspectives and constraints. *European Journal of Agronomy*, *29*(2-3), 59–71. doi:10.1016/j.eja.2008.05.005

Senanayake, R. (1991). Sustainable Agriculture: Definitions and parameters for measurement. *Journal of Sustainable Agriculture*, *1*(4), 7–28. doi:10.1300/J064v01n04_03

Sharma, L., & Mehta, N. (2012). Data Mining Techniques: A Tool for Knowledge Management System in Agriculture. *International Journal of Scientific and Technology Research*, *1*(5), 67–73.

Sharma, M., Sharma, S., & Singh, G. (2018). Performance Analysis of Statistical and Supervised Learning Techniques in Stock Data Mining. *Data (København)*, *3*(4), 54. doi:10.3390/data3040054

Sharma, M., Singh, G., & Singh, R. (2017). Stark Assessment of Lifestyle Based Human Disorders Using Data Mining Based Learning Techniques. *IRBM*, *38*(6), 305–324. doi:10.1016/j.irbm.2017.09.002

Shashi. (2009). Atmospheric Temperature Prediction using Support Vector Machines. *International Journal of Computer Theory and Engineering*, *1*(1).

Shaw, M. J., Subramaniam, C., Tan, G. W., & Welge, M. E. (2001). Knowledge management and data mining for marketing. *Decision Support Systems*, *31*(1), 127–137. doi:10.1016/S0167-9236(00)00123-8

Shivnath Ghosh, S. K. (2014). Machine Learning for Soil Fertility and Plant Nutrient Management. *International Journal on Recent and Innovation Trends in Computing*, *2*(2), 292–297.

Smart Farming Challenges. (n.d.). Retrieved May 15, 2019, from https://easternpeak.com/blog/iot-in-agriculture-5-technology-use-cases-for-smart-farming-and-4-challenges-to-consider/

Snehal, S. D. (2014). Agricultural Crop Yield Prediction Using Artificial. *International Journal of Innovative Research in Electrical, Electronic*, *1*(1).

Solutions for Smart Farming. (n.d.). Retrieved January 25, 2019, from https://www.kaaproject.org/agriculture/

Somvanshi, V. K. (2006). Modeling and prediction of rainfall using artificial neural network and ARIMA techniques. *J. Ind. Geophys. Union*, *10*(2), 141–151.

Sonka, S. (2016). Big Data: Fueling the Next Evolution of Agricultural Innovation. *Journal of Innovation Management*, *4*(1), 114–136. doi:10.24840/2183-0606_004.001_0008

Stockle, C. O., Martin, S. A., & Campbell, G. S. (1994). CropSyst, a cropping systems model: Independent Component Analysis. *Journal of Machine Learning Research*, *5*, 239–253.

Sundmaeker, H., Verdouw, C., Wolfert, S., & Pérez Freire, L. (2016). Internet of food and farm 2020. In O. Vermesan & P. Friess (Eds.), *Digitising the Industry - Internet of Things Connecting Physical, Digital and Virtual Worlds* (pp. 129–151). River Publishers.

Mehta & Kathiriya. (2015). Survey of Data Mining Techniques in Precision Agriculture. *International Journal of Standardization Research*.

Thakura, M., Olafssonb, S., Leeb, J. S., & Hurburgha, C. R. (2010). Data Mining for Recognizing Patterns in Foodborne Disease Outbreaks. *Journal of Food Engineering*, *97*(2), 213–227. doi:10.1016/j.jfoodeng.2009.10.012

Tripathi, S., Srinivas, V. V., & Nanjundiah, R. S. (2006). Downscaling of precipitation for climate change scenarios: A support vector machine approach. *Journal of Hydrology (Amsterdam)*, *330*(3-4), 621–640. doi:10.1016/j.jhydrol.2006.04.030

Tripathy, A. K., Adinarayana, J., & Sudharsan, D. (2009). Geospatial Data Mining for Agriculture Pest Management - a Framework. *17th International Conference on Geoinformatics*, 1-6. 10.1109/GEOINFORMATICS.2009.5293296

Tyagi, A. C. (2016). Towards a Second Green Revolution. *Irrigation and Drainage*, *65*(4), 388–389. doi:10.1002/ird.2076

Unay, D., Gosselin, B., Kleynen, O., Leemans, V., Destain, M. F., & Debeir, O. (2011). Automatic Grading of Bicolored Apples by Multispectral Machine Vision. *Computers and Electronics in Agriculture*, *75*(1), 204–212. doi:10.1016/j.compag.2010.11.006

Verheyen, K., Adriaens, D., Hermy, M., & Deckers, S. (2001). High-resolution continuous soil classification using morphological soil profile descriptions. *Geoderma*, *101*(3-4), 31–48. doi:10.1016/S0016-7061(00)00088-4

Vijaykumar, A., & Rajinikanth, T. (2013). Estimation of the Influence of Fertilizer Nutrients Consumption on the Wheat Crop yield in India- a Data mining Approach. *International Journal of Engineering and Advanced Technology*, *3*(2), 316–320.

Wang, Z. & Mujib, A. B. M. (2017). The Weather Forecast Using Data Mining Research Based on Cloud Computing. *Journal of Physics: Conference Series*, *910*, 012020. doi:10.1088/1742-6596/910/1/012020

Washington Okori, J. O. (2011). Machine Learning Classification Technique for Famine Prediction. *Proceedings of the World Congress on Engineering*, 2.

Weber, R. H., & Weber, R. (2010). *Internet of Things*. New York: Springer. doi:10.1007/978-3-642-11710-7

Weiss, G. (2009). *Data Mining in the Telecommunications Industry*. Academic Press.

Wolfert, J., Sørensen, G., & Goense, D. (2014). A Future Internet Collaboration Platform for Safe and Healthy Food from Farm to Fork. In *Global Conference (SRII)*, (pp. 266–273). IEEE. 10.1109/SRII.2014.47

Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M. (2017). Big data in smart farming [review]. *Agricultural Systems*, *153*, 69–80. doi:10.1016/j.agsy.2017.01.023

Xu, R., & Wunsch, D. (2005). Survey of clustering algorithms. *Neural Networks*. *IEEE Transactions on*, *16*(3), 645–678. PMID:15940994

Yadav, R. K., & Khatri, R. (2016). A Weather Forecasting Model using the Data Mining Technique. *International Journal of Computers and Applications*, *139*(14), 4–12. doi:10.5120/ijca2016908900

Zaki, M. J. (1999). Parallel and distributed association mining: A survey. *IEEE Concurrency*, *7*(4), 14–25. doi:10.1109/4434.806975

Zhang, N., Wang, M., & Wang, N. (2002). Precision agriculture—A worldwide overview. *Computers and Electronics in Agriculture*, *36*(2–3), 113–132. doi:10.1016/S0168-1699(02)00096-0

Zhang, S., Zhang, C., & Yang, Q. (2003). Data preparation for data mining. *Applied Artificial Intelligence: An International Journal*, *17*(5-6), 375–381. doi:10.1080/713827180

ADDITIONAL READING

Cambra Baseca, C., Sendra, S., Lloret, J., & Tomas, J. (2019). A Smart Decision System for Digital Farming. *Agronomy (Basel)*, *9*(5), 216. doi:10.3390/agronomy9050216

Pedersen, S. M., & Lind, K. M. (Eds.). (2017). *Precision Agriculture: Technology and Economic Perspectives*. Springer International Publishing AG. doi:10.1007/978-3-319-68715-5

Poonia, R. C. (2018). *Smart Farming Technologies for Sustainable Agricultural development*. Advances in Environmental Engineering and Green Technologies IGI Global.

KEY TERMS AND DEFINITIONS

Application Programming Interfaces (APIs): The code with the protocols that govern the access point(s) for the server and allows applications to communicate with one another.

Artificial Intelligence (AI): AI is a field of study which tries to make computers "smart" and the ability of a computer program or a machine to think and learn.

Artificial Neural Network (ANN): A computational model similar to the structure and functions of biological neural networks.

Bayesian Networks (BN): A marked cyclic graph describes random variables and conditional dependencies.

Cloud Computing (CC): Storing and accessing data and programs in the remote server over the Internet, rather than a local server or a personal computer.

Decision Tree (DT): A graph illustrates every possible outcome of a decision using a branching method.

Fuzzy Logic: A computing method uses "degrees of truth" logic rather than the usual "true or false" Boolean logic.

Genetic Algorithms: A heuristic search method used in artificial intelligence and computing.

Global Positioning System (GPS): A satellite navigation system used to determine the ground position of an object.

Information and Communication Technologies (ICT): The technology that use telecommunications to provide access to the information.

Internet of Things (IoT): A network used to collect and exchange data among the internet connected objects.

Knowledge Discovery in Database (KDD): Refers to the broad process of discovering useful knowledge in data.

Macronutrients: Calcium, Magnesium, Sulphur.

Micronutrients: Nitrogen, Potassium, Phosphorous.

Nearest Neighbor (NN): A non-parametric method used for regression and classification.

Rough Sets: A new mathematical approach to imperfect knowledge and defined by two sets, the lower approximation and the upper approximation.

Rule-Based Classifiers: A sets of rules for classification and uses term absence and presence for the classification.

Support Vector Machine (SVM): It is a discriminative classifier formally defined by a separating hyperplane.

Chapter 8
Agro Guardian:
A Smart Agriculture Framework for Precision Farming

Sanjeevakumar M. Hatture
Basaveshwar Engineering College (Autonomous), India

Susen P. Naik
Basaveshwar Engineering College (Autonomous), India

ABSTRACT

The mechanization of the process creates agriculture-based jobs for farmers, providing financial support and facilitating affordable agriculture equipment and machineries. Fruits markets are subject of opportunity and it is important to the suppliers to identify the quality of fruits based on the ripeness level of fruits before selling out in order to get higher level of profit. The proposed framework is an Android application in native language of the farmer to help the jobless farmers to find agriculture-based jobs suitable to their skill set and receive investments from various investors across the country. Further, it finds investment for the needy farmers and create suitable agricultural employment for jobless farmers so that there is an increase in the progress in the field of agriculture. It also facilitates the farmers with advanced equipment for performing various agricultural tasks, obtains the land on lease, and determines various stages of ripeness of fruit and provides the information about the government project and funding facilities.

INTRODUCTION

Agriculture is acceptably in which people have begun to be lethargic, forgetting so far what is keeping them energetic and alive. Although there are many numbers of hardworking, obsessive farmers and their life runs or rests only on farming or agriculture. However, there's is lot of corruption keep's on increasing these days due the intervention of the third party in marketing the agriculture products. The main motivation behind Agricultural Marketing and Department of Agricultural Business is to ensure feasible price

DOI: 10.4018/978-1-5225-9632-5.ch008

to the agriculture product and helps the farmer in competitive marketing scheme and also in implementing modern technologies that reduces the losses to farmer and encourages in more cultivation. India is most populated country, as compared to different economic category. Agriculture is important sector for economic in many countries like India. Presently the agriculture area contributes 18% of Gross Domestic Production (GDP). Various private and government agencies have straightly involved in agriculture area for develop the economics of the country. There is no success is obtained in agriculture sectors due to poor agriculture price, difficulties in contacting the dealers, lack of technology information and network connectivity etc. No proper mechanism or system to alleviate these complications.

The fundamental desire of farming is to bring improved marketing facilities, reduce trade performance, and eliminate market expenditure which facilitates the prominent budget to the efforts made by the farmers. Farmer is one of the most important people in every human being's life since the beginning of civilization. India gains food only if farmer cultivates the crops and manages many agriculture activities. There is much responsibility to farmer which makes their life tough and difficult. Even though farmer is rich enough to serve the food for every individual of the society but poor himself due to improper pricing for their cultivation, illiteracy, lack of the global knowledge, dependency on single source. Farmer remains happy if the cultivated crops are in good quality by the favorable environmental condition and better pricing. Since, the farmer's family is depending on cultivation if crops are not as per the demands then Farmer life becomes depressed. Farming work is a low admired job, especially regular agricultural workers. In this modern era industry and construction sectors have occupied majority of agriculture's labors due to the high wages paid. Most of the times small scales farmer depend on the external sources like relatives, companions, friends and others and they end up in giving high interest rates (36 to 120% per year) for requirements that might be in the form of money. Despite of the large machineries and modern technology involved in agriculture some parts of the country agricultural operations and mechanisms are carried out by human hands or labor using simple and traditional tools for implementing wooden, plough, seeds sowing etc. It results in huge investments of money on human labor. Land is the most valuable property in Framer life. For many new farmers, especially in areas where lands are quite expensive, it becomes very difficult to afford for their own the land to farmers and searching land for cultivation becomes complicated. The entire factor provides bad impact on the economic situation.

Growths of local fruit or vegetables industries in India are very immense. But, exporting fruits to other is country is less compared to other products. Fruits market is always option for consumers. It is very essential for suppliers and farmers to identify and deal the quality of fruits before sending out to market. Currently, human beings are experts' in grade the agriculture products that is by examining on the visibility features. However, physical analysis provides the inaccurate, incompatible and ineffectual on denigrating the quality of farming products. The physical analysis after a certain period of times the farmer gets board of this job. In order to overcome this problem, there should be a automatic analysis for determining the quality of fruits which provide the results in more accurate and agreeable output. By these results, it is beneficial in saving time and manual labor which in turn helps in the development of economic.

In the 21th century all the human beings live in modern, digital and technical adapted for their daily requirements. The smart phone is most commonly used all over, and farmers have started to use mobile phone for communication. But they need to recognize the significance of smart phone and new technology, so it can help them in direct communication involving the farmers and buyer or customers. By investing and considering the entire problem faced by farmers the author has introduced an android application which involves various features. Farmers need to download this application on their mobile phones so that they can easily search for jobs online as well as post ads regarding the jobs most farmer now have to understand in operating and maintaining sophisticated machinery. So, it provides them with an opportunity to post their machineries for rent to other farmers and also help the other farmers to find machineries for their work. This application also helps the farmers in finding the investors who are ready to invest on lands also there is another feature which helps them in finding lands for lease. It helps them in selling and buying agriculture commodities and services locally without middlemen through a add/ listing which they can post right from their mobile. Finally, it helps the farmer to check fruit ripeness level easily at anyplace and anytime.

LITERATURE SURVEY

Many researchers developed the agriculture application by designing algorithm for agriculture job, marketing, finance and investment and detecting the ripeness level. Some of the techniques presented in the paper are summarized in the following:

The relationship between the fruit image collected from the different angles and sweetness level are explored in Fruitylicious (Iswari et al., 2017) The sweetness level of a fruit is measured using Brix degrees' units. The 'refract meter' is utilized to determine the sweetness level. Further K-Nearest Neighbor condition is engaged to match the fruit picture or image along with its sweetness level of three varieties of fruits namely banana, apple and melon. (Dale et al., 2015) describes the quality and condition of the tomato constructed on base dimension, size and level of ripeness. The Algorithm utilizes in determine the shape, color detection and dimension of tomato. Color detection done by the ripeness determination that is by edge detection algorithm. All the connections with Raspberry Pi kit, motors, conveyor belts, Pi cameras devices will be executed and makes cost efficient system for finding the size, structure and level of ripeness in tomato this method can be used in different fruit and vegetables to regulate the ripeness level.

The proposed a method is used for detect the soil condition by constructing a convenient device via Raspberry-Pi and various sensors used achieve parameters of air temperature, air humidity, soil moisture, and ground altitude. The effect shows whether the crops need to be planted in the farm or not, and also determine the co-ordinates the agriculture land and landmarks of those (Junfitrhana et al, 2016). The research contributes to the improvement of agricultural produce, especially for agriculturist in determining the types of plants which are suitable for planting. The author describes It use agro-ecological calendar to map agricultural land use. Phonological curves supported on NDVI and NIR were extracted (Gerald Forkuor et al, 2012). A decision tree classification supported on the curves has mapping of six broad classes. Results indicate the successful discrimination of irrigated land from other land use classes, The synergetic analysis of data from these satellite systems with contemporary field data on cropping cycles promises to improve agricultural land us mapping in this savanna region.

A mobile supported application for agriculturist which would help them in their farming activities (Singhal et al, 2011). It proposes an android based mobile application – Krishi Ville which would take care of the updates of the various agricultural commodities, weather forecast updates, agricultural news updates. The agricultural information price is collected, processed and disseminated to public through agricultural market information system research developed an Android version of KomoditiAceh by utilizing RESTful Web service (Mutiawani et al., 2017). The application can be utilized to view all commodities prices, view commodities prices by district, view price comparison graph and search or browse commodities. Internet of things (IOT) have a variety of application in precision farming agriculture sectors example examining the crop development, choice of fertilizer, irrigations judgment support system, etc. (Rajeswari et al., 2017) IoT equipments is used to capture agricultural data and stored it into the Cloud database. Cloud supports big data, Big data is used to analysis the data viz. fertilizer requirements, analysis of crops, market and stock requirements for crop. Later the forecast is performed by data mining methods in which the processed information should reach farmer through mobile application.

It is imperative to find an efficient solution for the agriculturist which can help them in cultivating crops which can give them a better yield and are suitable as per the present weather conditions. (Guptaa and Trivedi, 2016) It proposes an android based application, *e-krishakMitra* intended to address this issue. *e-krishakMitra* is a cloud-based application which provides both Hindi and English interfaces to seek various queries in real time related with the crop cultivation. The unique features of *e-krishakMitra* are its simple and user-friendly interface. It is approach is made to design and fabricate a closed-loop control system for soil-moisture monitoring and controlling. It is achieved through deployment of proposed Wireless Sensor Network (WSN) model in agriculture field. They have designed new sensor nodes with the help of Arduino Platform and Raspberry Pi module to minimize the cost of entire product (Misra, 2017). Selection of hardware component, firmware development, integration and field implementation, testing and optimization are the processes which can been followed to develop the product.

The application helps agriculturist can be connected to the end users and supply the agriculture product directly to them. So, the famer can be benefitted with the profit and consumers also get the quality product in affordable price from farmers (Khodaskar, 2015). A web service is method on Web server that a client application can make request from HTTP through web. ASP.NET which is able to make web services or built in application services and client application can call these services, web services is able to call the database functionalities like store and retrieve the data from database. Its empirical model of how the Internet of things can be applied to Indian agriculture. Proposed a model outline of how the IoT concept can be illustrated with respect to Agricultural practices (Gayatri, 2015). ZigBee is an organization which provides the standardization of Internet of Things manufacturing by only if an authoritative standard to the things utilized in Wireless Sensor Networks

Kisan Suvidha mobile app provides information on five critical parameters-weather, input dealers, market price, plant protection and expert advisories (Welfare). An additional tab directly connects the farmer with the Kisan Call Centre where agriculture experts answer their queries. Unique features like weather alerts and market prices of commodity in nearest area and the maximum price in state as well as India have been added to give power to agriculturist in the best possible manner. Technological importance has been a great support for making decisions in various fields especially in agriculture (Thankachan and Kirubakaran, 2014). The main aim is to provide agriculturist for their awareness, usage and perception in e-Agriculture. The results obtained indicated the stage of alertness is less such that there is a need for e-agriculture for their support. e-Agriculture is a platform for supporting marketing of agricultural products KrishiSense is a semantically aware web enabled wireless sensing system for precision agriculture applications. Through integration of Open Geospatial Consortium (OGC) specified Sensor Web Enablement (SWE) standards on sensing system has enabled the interoperability between various standardized sensing systems (Sawant et al., 2014). KrishiSense acts an interconnection between various users (researchers / scientists, agriculturist and extension community) through different protocols and distributed web connected platforms, thus facilitating human participatory sensing. iFarm, is proposed, to support efficient farming management. The system consists of Smartphone applications, Web browsers and a cloud server (Murakami, 2014). Cultivators on farmland can easily refer to work plans, enter field data into the cloud system, and share them with head office in real time by using smart phones. Farmers at office can analyze data in the cloud system with a Web browser and approximation farming costs and form work plans based on their analyses.

One of the most interesting fields having an increasing need of decision support systems is precision agriculture (PA). It presents WSN as the best way to solve the agricultural problems related to farming resources optimization, decision making support, and land monitoring. This approach provides real-time information about the lands and crops that will help farmers make right decisions (Kassim, 2014). Implementation of WSN in PA will optimize the usage of water fertilizer and also maximized the yield of the crops. Agro-App is mobile app constructed to benefit the cultivators and ordinary man those who needs to cultivate vegetables for their daily needs. It gives the farmer more updated providing all the information with respect to crop, pesticides, insecticides, financial sector etc (Aggarwal et al., 2014). It gives complete information for the famer which crop to cultivate in which seasons and crop suited for particular land or area it is mainly helpful for cultivators.

AgriCom provides farmer, buyer, seller and instructor who are representing cultivators four types of agents, buyers, sellers and technical instructors in agricultural community. The system uses common MySQL database as ontology and the common message space of the agents (Jayarathna & Hettige, 2013). Single agent in the multi-agent system is supported by a java application which is able to be worked as one of the above four types of persons such as cultivators, buyers etc. The system is developed using Service Oriented Architecture (SOA) to process knowledge base and spatial data. The system is an effort to fill the gap between cultivators and agricultural experts (Kumar et al., 2013). A farmer can give inputs connected to crops being cultivated and location specific information to get specific suggestions, alerts and recommendations to improve productivity. Whenever a farmer observes some abnormal behavior for crops or climate, the system is able to generate recommendations supported by inputs provided. The results are displayed on an Android based mobile devices for demonstration of the system.

It introduced the new terminology Agriculture Intelligence for agricultural business. (Ghadiyalim et al., 2011) has shown the current trends of the Agri-Business concept using Intelligence, limitation of current Information Technology in the field of agriculture and tries to overcome the same. Authors have also proposed architecture model. Such architecture takes care of data availability from a range of resources and provides solution by applying agriculture intelligence. e-Agriculture involves the conceptualization, design, development, evaluation and application of pioneering ways to use information and communication technologies (ICTs) in the rural domain, with a primary focus on agriculture. e-Agriculture is an emerging field focused on the enhancement of agricultural and rural development through improved information and communication processes (Chandra & Malaya, 2011). The author discussed the problems & prospects of e-Agriculture in Rural Development in Indian context. Agriculture is one of the dominant fields that shape the socioeconomic development of any country (Mohamad et al., 2017). Technological advancements and innovations served as utensils to distribute knowledge and practices of agricultural activities and make improves lives for cultivators, traders, policy makers, and the overall society. It is obvious that Knowledge has become a very significant factor in production, food security, education, poverty alleviation, and other millennium development goals. Further the hyperspectral imagery (Karchi & Nagarajan, 2018, 2017; Rashmi et al,2018) can be employed to monitor the growth of the crop, vegetation etc.

Summary of the Literature Survey

Various paper uses different techniques for e-commerce agriculture the algorithm utilizes are such as USB GPS programming, Application Programming Interface (API) Web Services, USE tool Usefulnes, Satisfaction and Ease, Classification methods Association rule algorithm, Programming Language, Java, Decision Support System, classification system supported by decision based tree . Determine the ripeness level it uses the algorithm such as k-Nearest Neighbour algorithm, Color and texture Feature Extraction, tomato Classification and Pre-processing.

Top Five Existing Android Agriculture Application

- **AgriApp:** It is the online marketplace for agriculture products, it also provides governments scheme provide to farmers and provide chat option with agriculture expert. It contains agriculture related videos.
- **Iffco Kisan:** This application makes use of memory with good interface. It provides information of agriculture advice, comparison of prices and farming tips. It alerts with weather forecast tips. The application contains 10 different languages across India.
- **Agri Media Video:** Its popular application for video category. It is online marketplace for agriculture products, it is provide with chat option to speak with expert to solve the problem by upload the images. It provided with agriculture news, Videos, new government schemes etc.
- **FarmBee:** This application makes use of memory with good interface. The application contains 10 different languages across India. It provides information about fertilizers and life cycle of crops at different stages. It provides the forecast information based on the user location
- **Kisan Yojana:** It is main built for providing the information about schemes for farmers it eliminates the gap between the farmer and governments. It basically saves the time for farmer to travel to government office

Importance's of Technology to Farmers

In advancements of Digital India, there is so many schemes provided to farmers where farmers are provided with mobile phones, tablets and electronic equipment so they can improve the farming activity. Creating Interactive programs in rural places and show casing the importance's and necessity of technology to farmers. For example, hosting radio station towards educating farmers on a new and modern technique, through television advertising and providing them the audio and video demonstration of the use of application and its importance.

Through Smartphone farmers can check the prices of various agriculture products with comparison so they can sell their agriculture products in better prices. By this the middleman communication can be completely reduced. Farmer can easily be connected to buyers and sellers through the common platform. Mobiles phones helps the farmers reduces the time and cost burden. Mobile phone helps to raise incomes and online marketplaces

Issues and Challenges Related to Agriculture

- **Need of Knowledge and High Percentage of Illiteracy:** Although approximately all farmers need to deal with the android phones and application need to be accessed through phones, just some knows how to utilize the phones they just know how to make and receive calls. Most of them unable read/ write and they cannot understand the messages and sending back is difficult as a result a number of have to be depend on educated companion, kids or others to guide them.
- **Apathy Towards New Technology:** The ability of technology is still extremely poor and large number of farmers are still unaware regarding such advancements. The distribution of technologies is not consistent in whole country. Farmers of rich states are always at receiving end like-Punjab, Haryana, Maharashtra and farmer still uses their old techniques in backward states
- **Lack of Financial Help:** Poor economic support from organization that allows farmers grows, increase, and keep up their yields. There are various Micro finance groups functioning in India presently, most of the farmers do not have access to these groups and many farmers may not even know by what means these groups work and how these groups can help them. The maximum numbers of farmers in India have poor financially practices it approximately impossible for farmer to approve new farming techniques.
- **Lack of Marketplace:** Marketplace for farmers is one of the biggest concerns for India presently which scatters the lives and living grades of millions of people. The major issue is how to generate markets and how to make them reasonable and most importantly, what make them grow? Farmer in numerous places grows only one type of crop. Every person at India consumes it and everyone grows it. But the disappointing thing is each one is not able to sell due to the deficiency of marketing facilities. Small scales farmer is impossible to get marketplace their agriculture yields. With Improved market availability can serve the poor famers with profit.

Issue and Challenges Related to Mobile Device

- **Network Connectivity in Interior Areas:** Various remote locations across the India is robust, stable internet or network connectivity is not accessible. if there is no network performance and bandwidth velocity then implementation of e-farming will stay challenging. Because several agro-sensors/gateways directly depend on cloud services for data sending and receive. In farmlands there are tall, huge trees or a hilly terrain, GPS signals becomes a large problem.
- **Battery Power of Android Mobile:** Having a short-lived battery power can restrict the moment of time spend in the farm or field. It may also be a safety concern if a person is lost. It's good to have an Android device but if it has no good battery power then it is not useful – even though the most recent location can be still be examined. It is not issue if there is alternative or back up battery or power sources.
- **Cost:** The cost of many android devices has scale down, but high-quality android mobile devices are still costly. The high-quality mobile devices will have better accurate and good mapping quality. However, there are many devices which are cheaper, device which uses low quality mapping which provide inaccurate results while using the devices.

Motivation and Problem Formulation

- Farmer is a human being engaged in agriculture, cultivating living creatures for food or raw resources. Farmer might have their own farming land or might work as laborer in different land owned by other people. Farmers usually start work early, and during planting and harvesting season work until dusk. His/her life can be made a little better if he can find all his requirements easily in his mobile phone, so this application is developed
- Design and develop an android application that have native languages and helps the farmers in finding jobs, buying/selling of their yield, finding affordable equipment, finding investment on their land and thus finding a land on lease and to determine the ripening level of fruits and vegetables.

Scope and Objectives

Agriculture is the backbone of Indian economy. Farmer plays a vital role in changing the agricultural landscape for the betterment of the country. Hence it makes a farmer's life easy by providing him with many online features he can contribute to the growth of agriculture.

The objective of the system is:

- The main foundation is the application has the Regional languages.
- Its purpose is to automate the process of creating jobs based on agriculture for the farmers.
- Providing financial support as well as providing affordable agriculture equipment and machineries.
- The farmers can be benefitted by the App as it will help the jobless farmers to find agriculture-based jobs suitable to their skill set and receive investments from various investors across the country.
- The main objective is to find investment for the needy farmers and create suitable employment for jobless farmers so that they increase the progress in the field of agriculture.
- Farmer also should be benefitted by ripe track that it helps the farmer to determine the level of ripeness of fruits and vegetables.

Proposed System

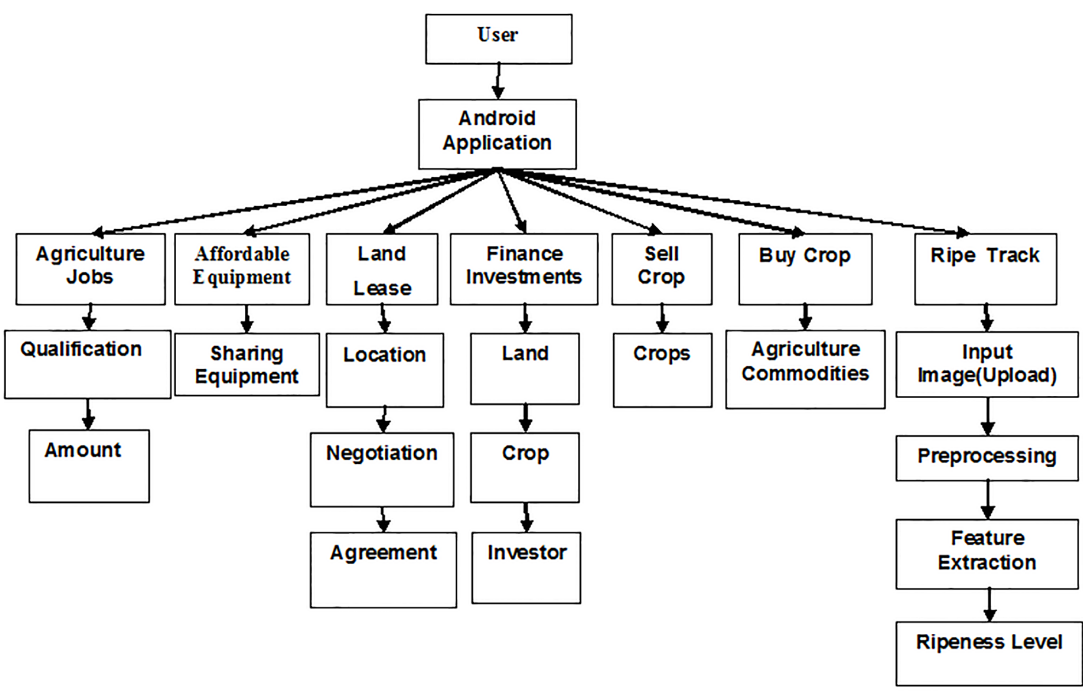
The framework of the proposed mobile application system to alleviate the problems associated with farmers in India is depicted in Figure 1. The framework has seven functionalities namely agriculture-based jobs, sharing of agricultural equipment, finding fields for lease, finding investment, selling crops, buying crops and tracking the ripening status of the fruit.

Regional Language

There is huge number of applications built for farmer which doesn't consist of their regional language. Agro guardian has language convertor which helps the farmer to understand the application easily.

Figure 1. Framework for agro-guardian

Figure 2. Block diagram of agro guardian



Agriculture Bases Jobs

As we all are aware about the difficulties that the farmers face in finding jobs, this android app makes it easier for the farmers to find jobs according to their qualification. As we have provided an option to post ads about the jobs available in those particular fields. This makes it easier for farmers so that they can directly apply for their interested jobs. As a result, we can overcome the unemployment problems faced by many people in India thus help in increasing the annual income of our country.

Sharing of Agricultural Equipment

Agro guardian provides an opportunity to the farmers to post ads about the equipment they wish to provide on rent to other farmers with some liable value. This helps the farmers who are in search of equipment for their farming obtain them with a reasonable amount of money. This facility helps the farmers in doing their works faster hence resulting in financial betterment of the country. As well reduces the amount of hard work that the farmers do in giving us daily crops.

Land Lease

Farmland is frequently leased for various reasons, including crop production. The process of locating, negotiating and agreeing upon lease terms can differ depending on factors such as location, size of the land plot and intended use.

Finance and Investments

This feature lets the customer own a farmland and also enjoy profits without incurring any maintenance costs. This feature helps in increasing incomes of small and marginal farmers through aggregation and development of agribusiness. It is provided with an option to upload the images of their particular land and also identify the crop cultivated there so that the investors can easily see the image and decide as to how much can he invest on this particular land.

Sell Crops

Agro Guardian is an android application for online agricultural marketplace, where a farmer can sell his crop, products online with competitive prices in India. This is beneficiary to farmers as it does not involve any mediators. This marketplace helps agricultural and rural-based businesses to expand market share through an online trading community. Agro guardian provides a facility for farmers who want to showcase their produce online, and traders can quote price from anywhere. It will result in increased number of traders and greater competition. It will also make sure open price discovery and improved returns to farmers.

Buy Crops

Agro Guardian is an android application it is designed to create a unified market for agricultural commodities. Buyer can directly purchase products from farmers using this android application, it benefits both buyer and farmers to facilitate convenient market to buy and sell all agriculture related products. In the application buyer can check all details of crops like crop quantity, offered price, quality of the product with the picture of product. After checking all details, a buyer can directly contact a farmer using his phone number provided in the database and buy the products of desired quantity.

Ripe Track

This feature aims to keep track of ripening status of the fruit in organic farming in order to avoid early plucking of the fruits from the tree. The process involved in the determination of fruits ripeness level without cutting or disturbing the fruit is more feasible and quicker. The image processing/sensor technique is employed to determine the ripening level of the fruit as the farmer get better price. The ripening level of the fruit can be efficiently recognized by employing the soft-computing techniques (Angadi & Hatture, 2011, 2018) and symbolic data analysis (Angadi & Hatture, 2019).

Example

Consider a person who has some land available for farming or needs a farmer to work in that land, then that person can post an ad specifying the need of farmer or regarding lending the land to any farmer for lease. He can even post an ad to rent any of the agricultural equipment or invest in any of the farmer's land to get profit on the yield generated.

Scenario (if the farmer uses the application):

- The farmer can apply only for the suitable agricultural jobs posted.
- The farmer can borrow the equipment or machineries on affordable price instead of buying brand new equipment which may cost much more.
- The farmer can search and reach out to the investors who are ready to invest
- The farmer can scan the fruit image and determine the ripeness level of it.

Overview of Development Environment

Various tools and technologies have been utilized to develop Agro Guardian. As the

Application is supported on Android it's been developed using Android Studio IDE and Python as Backend Programming Language.

System Application

Android holds arrangements of core application in supports of electronic mail, messaging, date and time with calendars, accessing, contact, and many more. Application holds platforms which contain application with no special conditions where users choose to install. But the third-party application happens to be users default web browser, SMS messenger, or even the default keyboard. It provides Functionality for both for users are provided with basic functionalities and developers have authority to access from their own application

Android Studio

Android Studio are the standard integrated development environment (IDE) for Android platform. Established on JetBrains' IntelliJ IDEA software, Android Studio is structured exclusively only for Android progression and replace with Eclipse Android Development Tools (ADT) as Google's chief IDE for native Android app development.

Android Studio contributes the fastest tool used for constructing application on each types of Android device. World-class coding and editing, complication and debugging, performance and error checking tooling, compatible build system, and instantaneously in build/deploy system and provides privilege to focus on constructing different and high graded quality application.

Python Tools

Python is used for high-level program language used in general purpose programming, established by Guido van Rossum which was released in 1991. Python is designed for code readability and provides syntax which helps the programmers to express in just few lines of code rather than using C++ or java languages. The python has dynamic system and memory managements and integrates different programming languages with object oriented and it has wide range of library packages

Flask Framework

Flask is a Python web framework in addition to it is light weighted and supported on Werkzeug toolkit and Jinja2 template. It is Berkeley software distribution licensed. It always aims to keep the core of application simple and provide extensible. Flask is unique that is self-labeled as "micro framework" since it does not need any particular tools or libraries. Flask does not have any built-in abstraction layer for database, do not have form a validation supports. Instead Flask supports the extensions to add such as functionality to the application. Extensions be present for object-relational mappers, form a validation support, and upload handling, several open authentication technologies and familiar framework related tools. Extensions provide updates more frequently than that of core Flask program.

Anaconda

Anaconda is the most important open data science platform powered by Python. The open source version of Anaconda is a high-performance distribution of Python and R and includes over 100 of the mainly popular Python, R and Scala packages for data science. Additionally, they have access to over 720 packages which can be easily installed with anaconda, renowned package, dependency and environment manager to facilitate is included in Anaconda.

JSON

JSON (JavaScript Object Notation) is open standard format that uses easily readable text to transmit data objects consisting of key-value pair. It is common data format used for asynchronous browser and server communication replacing XML. JSON is cross language data format. JSON is compatible with most of languages available in market and JSON is mainly used for exchanging data. JSON is built with key and value pairs. Key is a string and values are numbers, strings, objects or Boolean.

SQLite

SQLite is software library that implements self-contained, serverless, zero-configuration, transactional SQL database engine. SQlite is most commonly deployed SQL database in the world. SQLite is a system for managing database. It is relational database management system with C programming library. The SQLite system is not client-server database management system and it consist of built-in library in which one can address to it or directly access it there is no need of additional installation and configuration process. Usually, SQLite database is maintained in computer disk in only single disk with all objects such as tables, data view, event handler and others it doesn't require any dedicated server. SQLite is necessary to save for file archive; it is also used to create small size archive and include larger amount of metadata than zip file.

Design and Implementation

Design

Designing an efficient layout for the Application is as important as developing the backend. Designing is to make a visual language for our users which synthesizes the classic principles of good design with the innovation and possibility of technology and science. Using Data Flow Diagrams for the efficiency of the processes involved in the applications is very important.

Data Flow Diagrams

The Figure 3 shows the flow diagram for the login and registration process of the application.

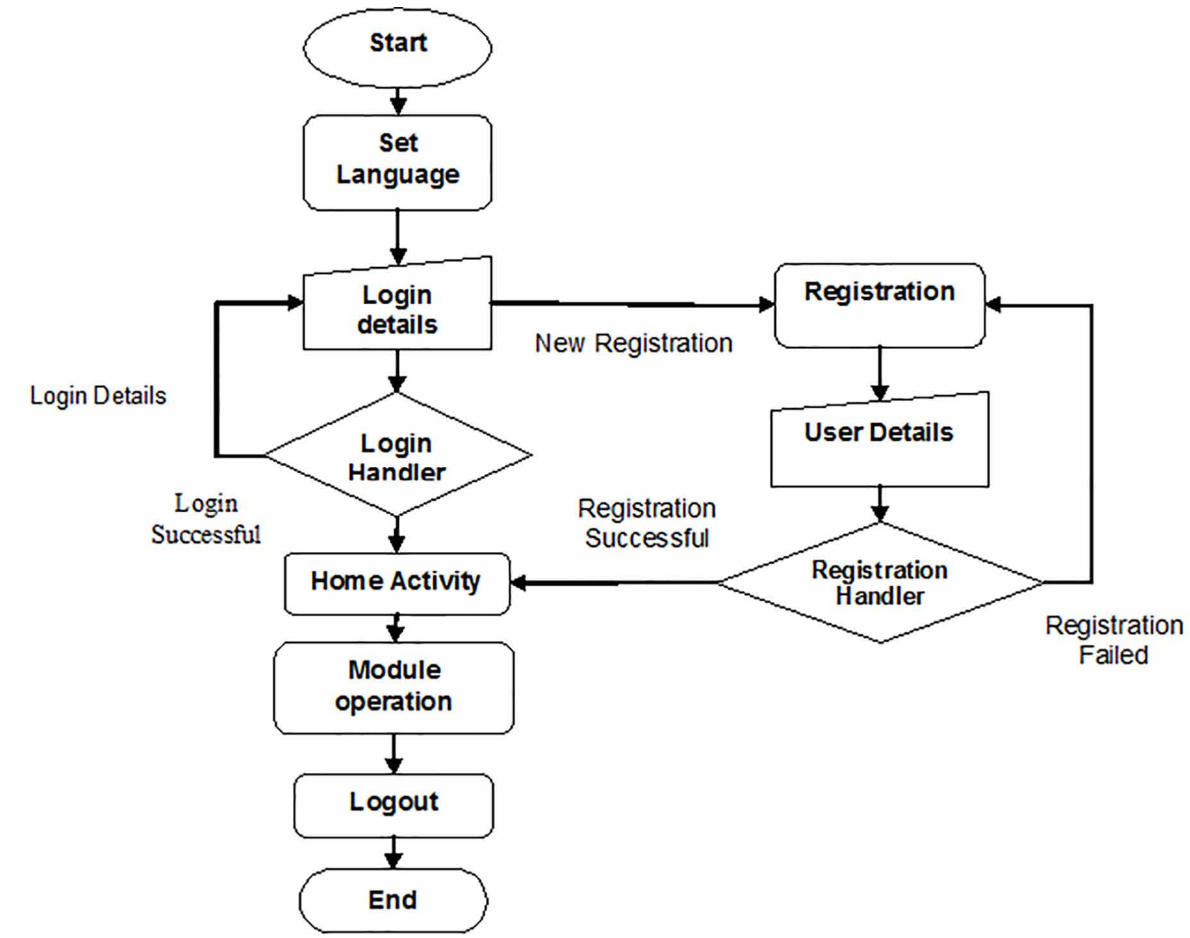
The Figure 4 shows the flow diagram for applying and requesting process in the application. The user needs to select one of the seven modules i.e. JOBS, SHARE EQUIPMENTS, LAND LEASE, FINANCE AND INVESTMENT, SELL CROPS, BUY CROPS AND RIPE TRACK. Then fill the details and post the ad and process the request according to the module.

Implementation

The Agro Guardian Application is developed using Android StudioIDE to create interactive Graphical User Interface and integrate the code which communicates with the server-side code.

The Figure 5 shows architecture design of the application. It follows the basic Client Server Architecture i.e. The User performs the operations through the Agro Guardian Application, the data request and response are done through JSON Parsing using Volley Library. At the server-side FLASK (Python Framework) is utilized for CRUD operations which follows' the MVC Architecture. SQLite3 is utilized as database. The application is developed for two languages i.e. English and Kannada.

Figure 3. DFD for login and registration activity



Implementation

Agriculture Bases Jobs

Farmer need to login to the application after the validation then he/she can create advertise(ad) which provided an option to post ads about the jobs available in those particular fields. They need to add the details of job according to their qualification. It consists of three fields in Figure 6.

Job Title: Name of the Job, Job Description is Information about the qualification they have in certain fields or experience they have and payment is price allotted for the job.

The details are stored in database and any other user can view the post and apply for the particular job based on his criteria defined.

Figure 4. DFD for applying / requesting process

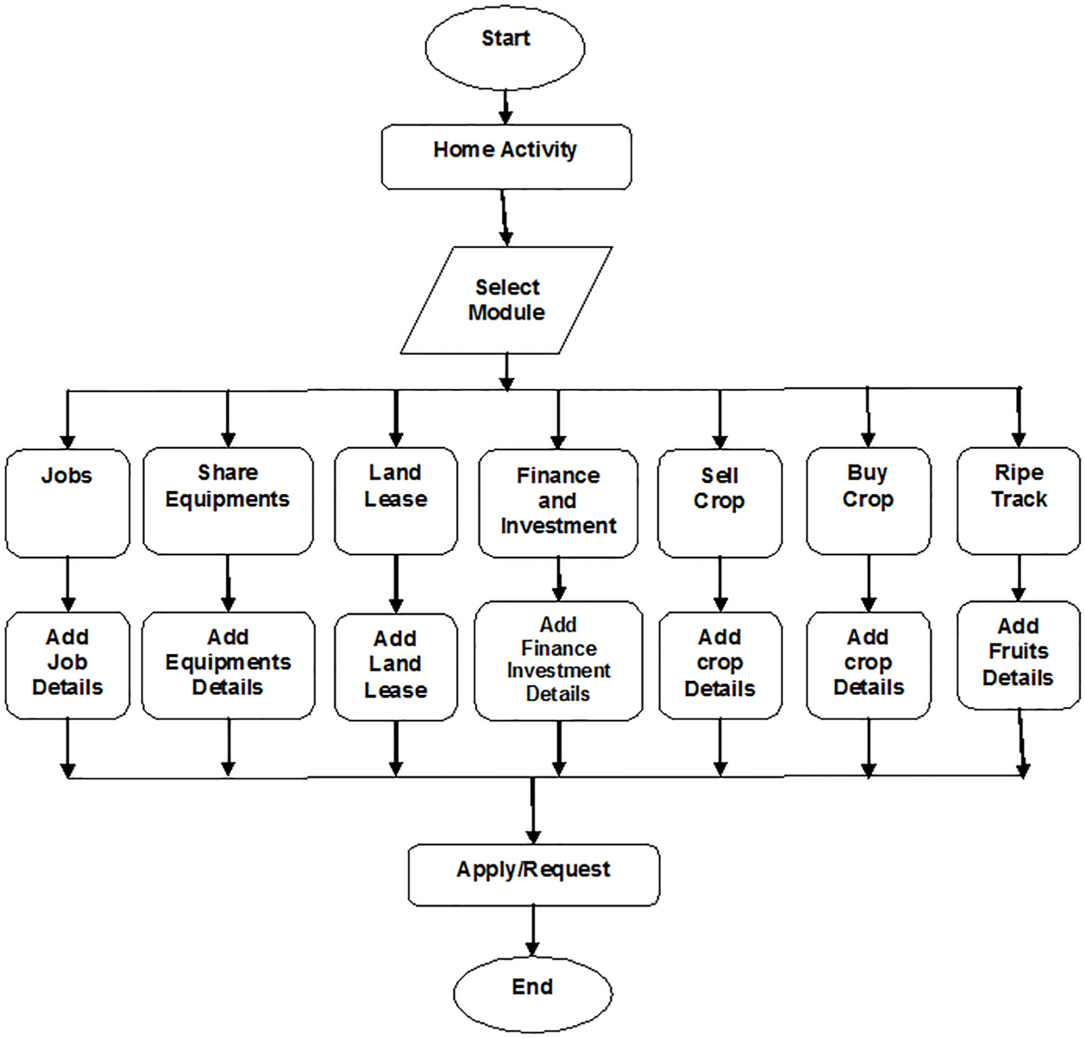


Figure 5. Agro guardian architecture design

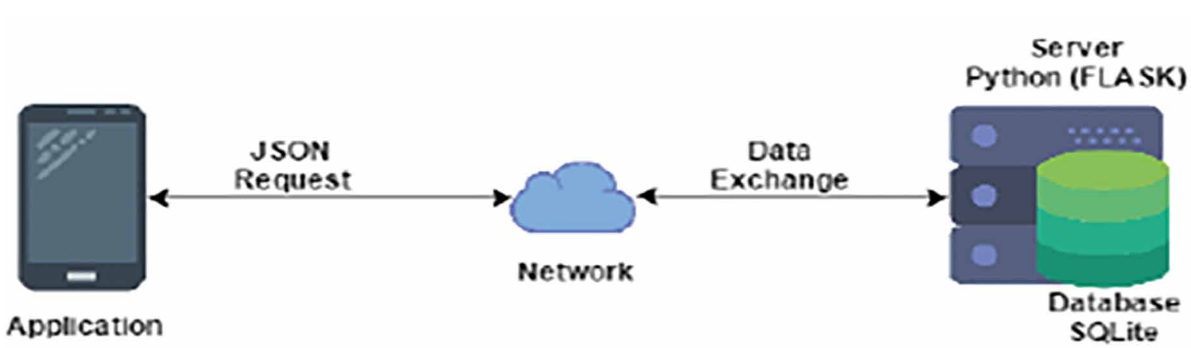
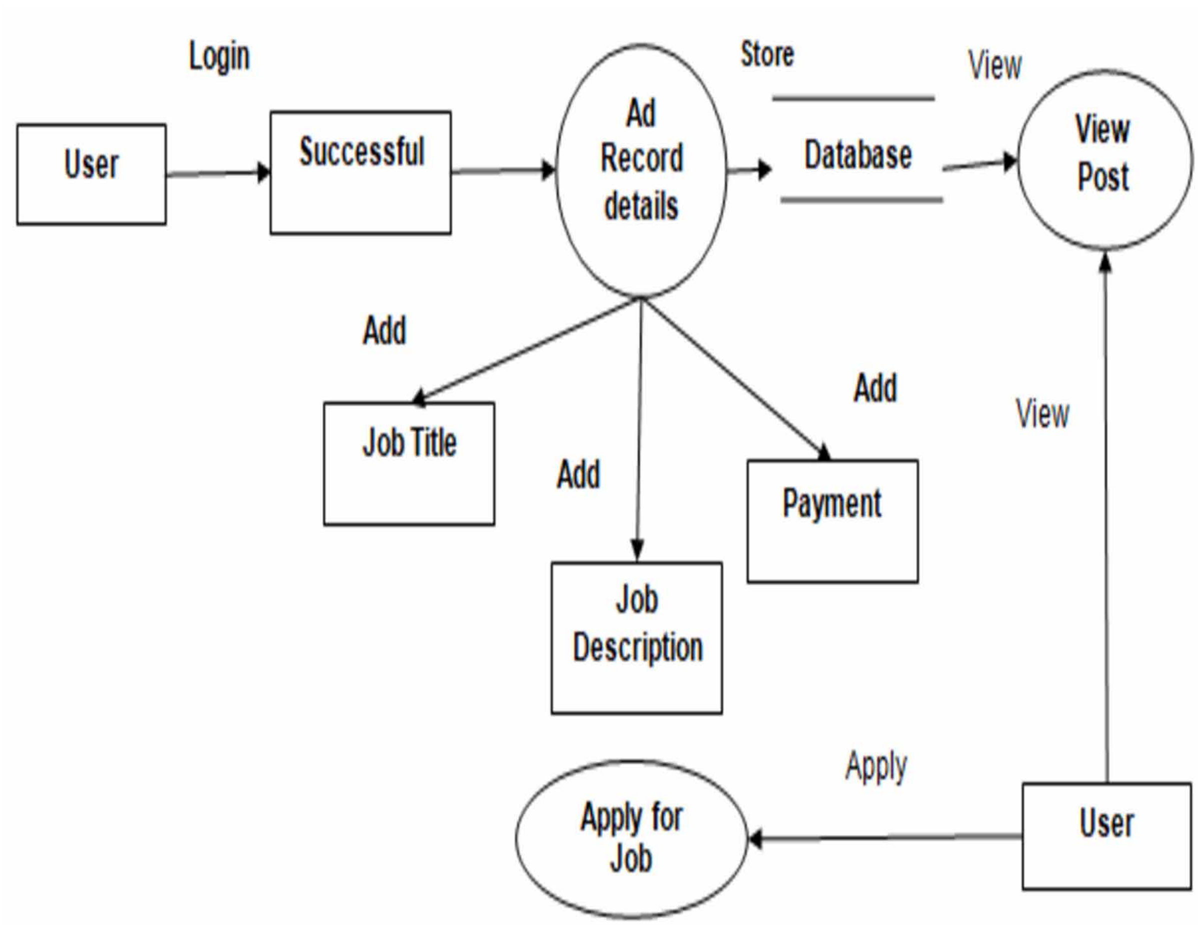


Figure 6. Data flow for agriculture jobs



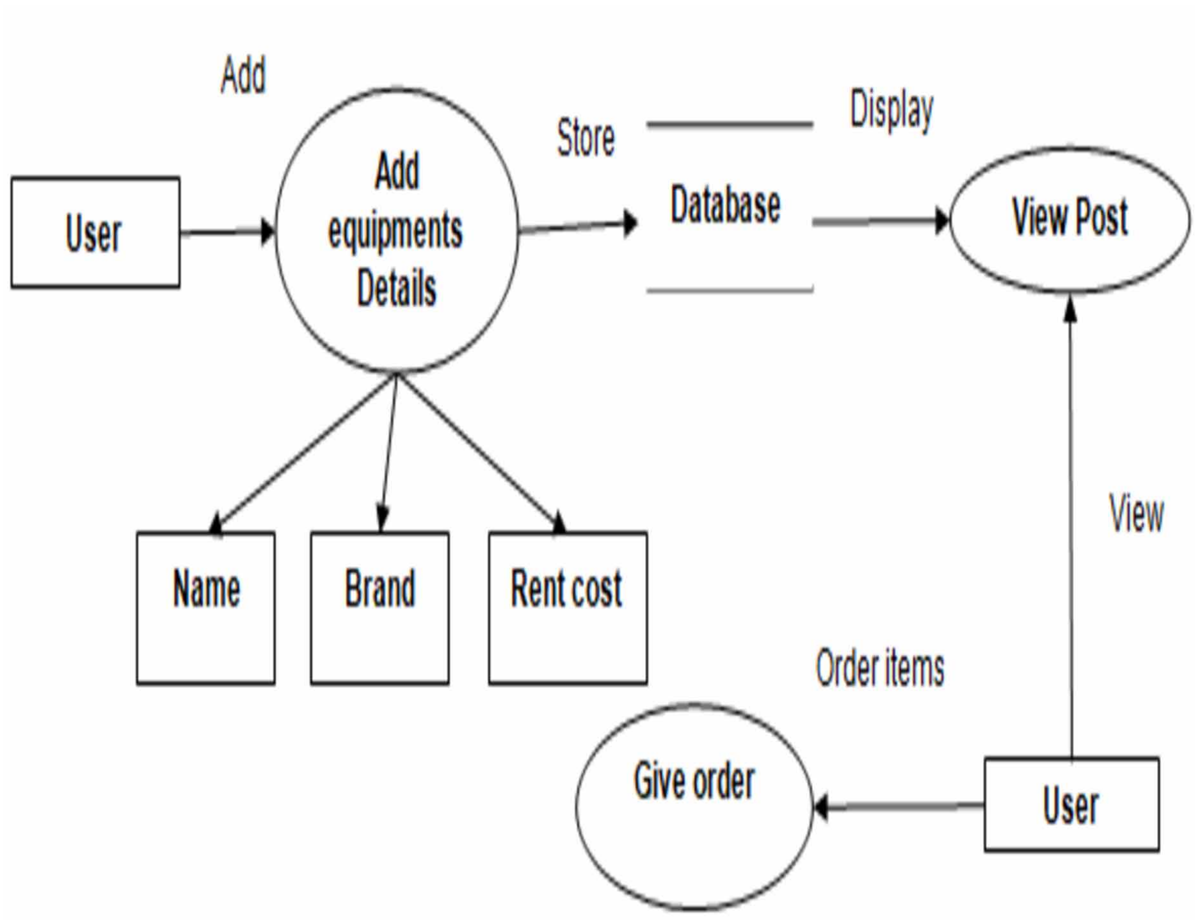
Affordable Equipment

The farmer can add the details of equipment such as Name of the equipment, brand name of Equipment to understand the quality of the Equipment and Rent Cost. The details are stored in the database automatically and user can assess it that is he/she can view it and place the order as represented in the Figure 7.

Land Lease

Farmer need to login to the application after the validation, he/she can add the land details, so that it can be provided for lease. It includes the entity such as adding the Land size, Land size is the measured size of land, soil type is given because the buyer can think about the cultivation of crop, landmark is the location of the land, years are the duration period for lease, cost is the price depending on the land size and years and finally the contact details of the farmer as represented in the Figure 8. All the necessary fields are stored in database. Other farmer can view the details of land if they wish for purchase of land.

Figure 7. Data flow diagram of affordable equipment



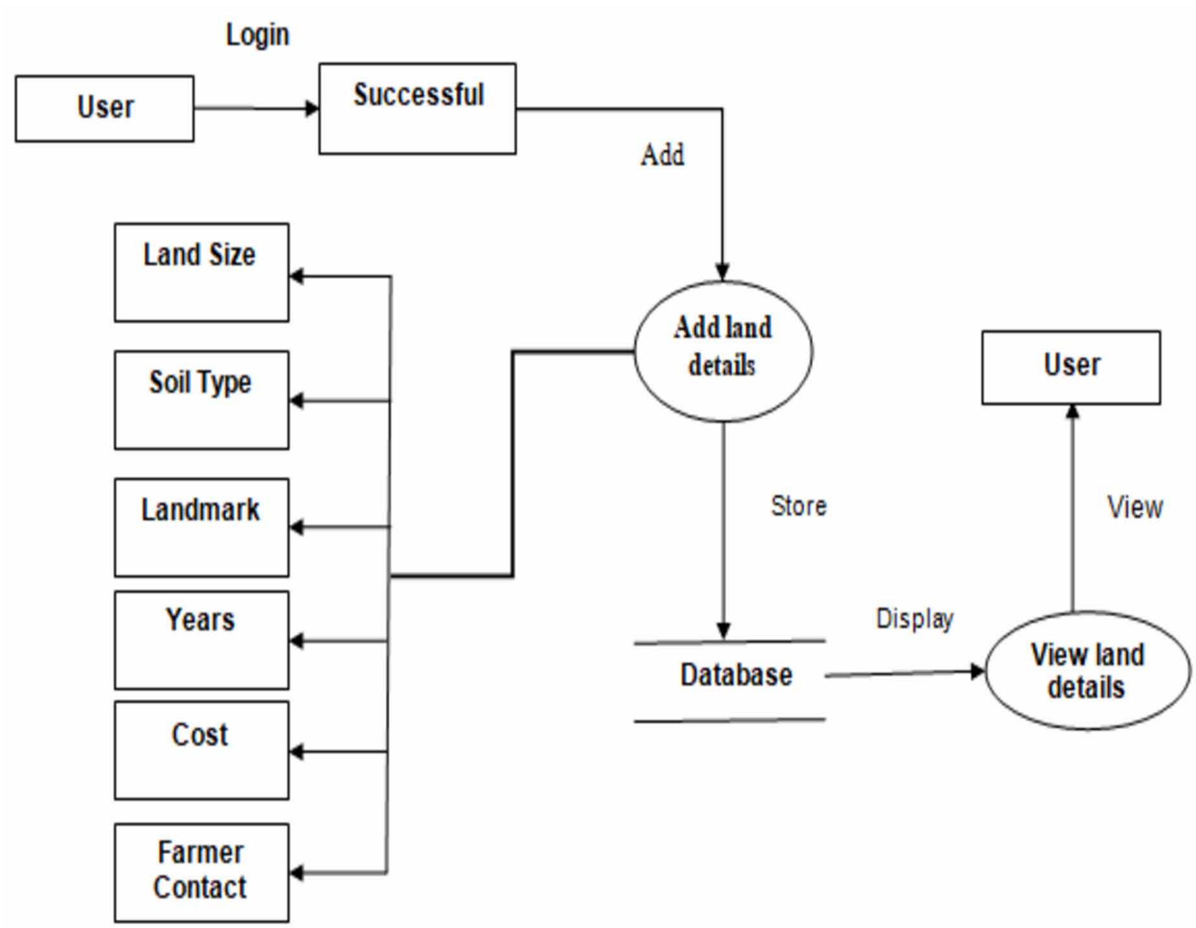
Finance and Investment

Land lease is connected to Finance and investment there is similar entity included in it. The major difference is the Investor need to add the details of investment in particular fields that depends on the land size, soil type, year, cost and landmark as shown in Figure 9. Details are stored in the database if any user finds are ready to invest they can apply for it

Sell Product

The Farmer stores the details of crop which they want to sell, such as Name of the crop, description of crops, price per kilogram and quantity. All the details are stored in the database other user can view it and place a order if they wish to purchase as shown in Figure 10.

Figure 8. Activity for land lease



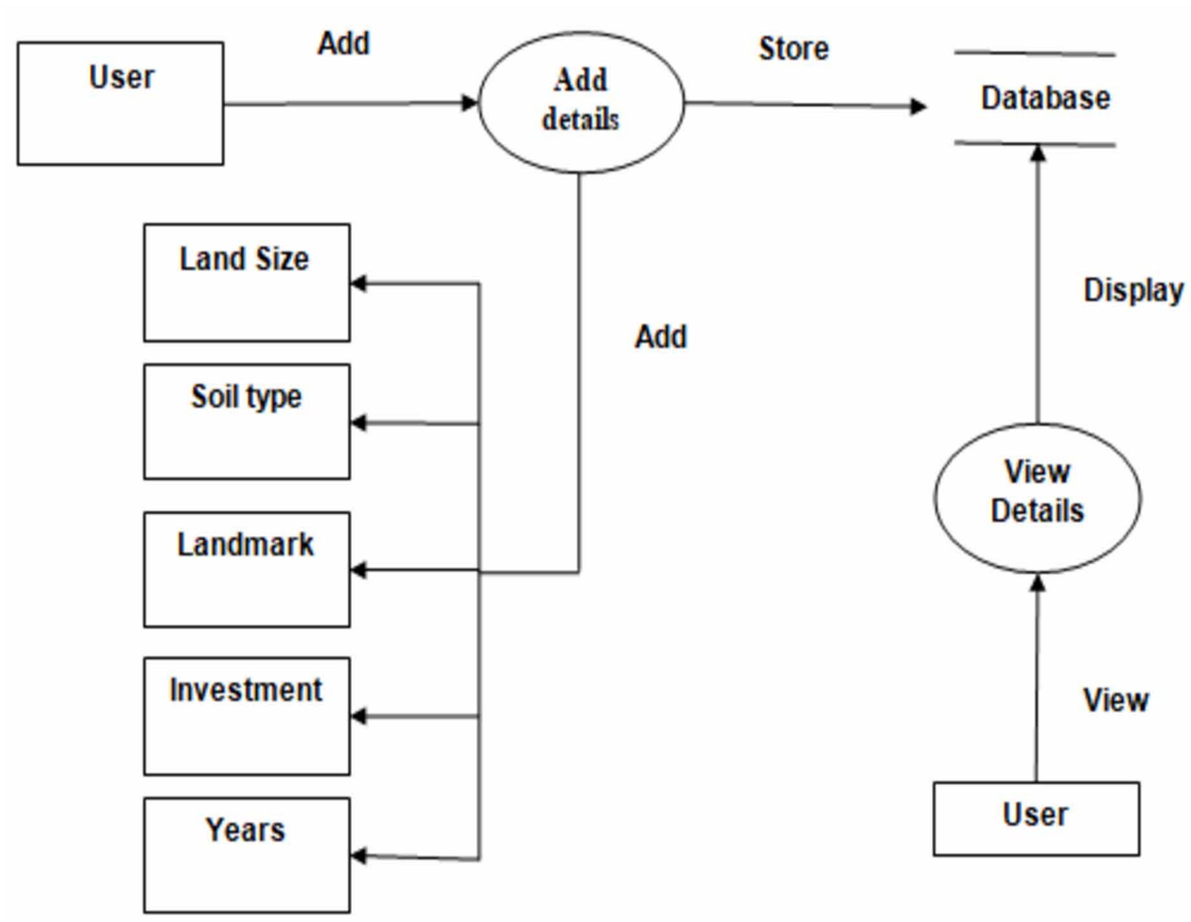
Purchase Product

If user wishes to purchase any product from the farmer, they give an order or request for crops. User need to add the details of request for crop such as Name of crop, Description, Quantity and it is stored in database another user can respond for the request as shown in Figure 11.

Ripe Track

The color and texture information of the fruit image is extracted and compared with the knowledgebase using k-Nearest Neighbour (KNN) classifier. The ripe track mobile application module is implemented and executed using image processing techniques and android. In the proposed work three varieties of fruit images were captured namely apple, banana and melon. The fruit images are captured from different positions with orientations. The sample fruit images are shown in the Figure 12. Method used for the feature extraction on each individual fruit is described below.

Figure 9. Activities for finance and investment

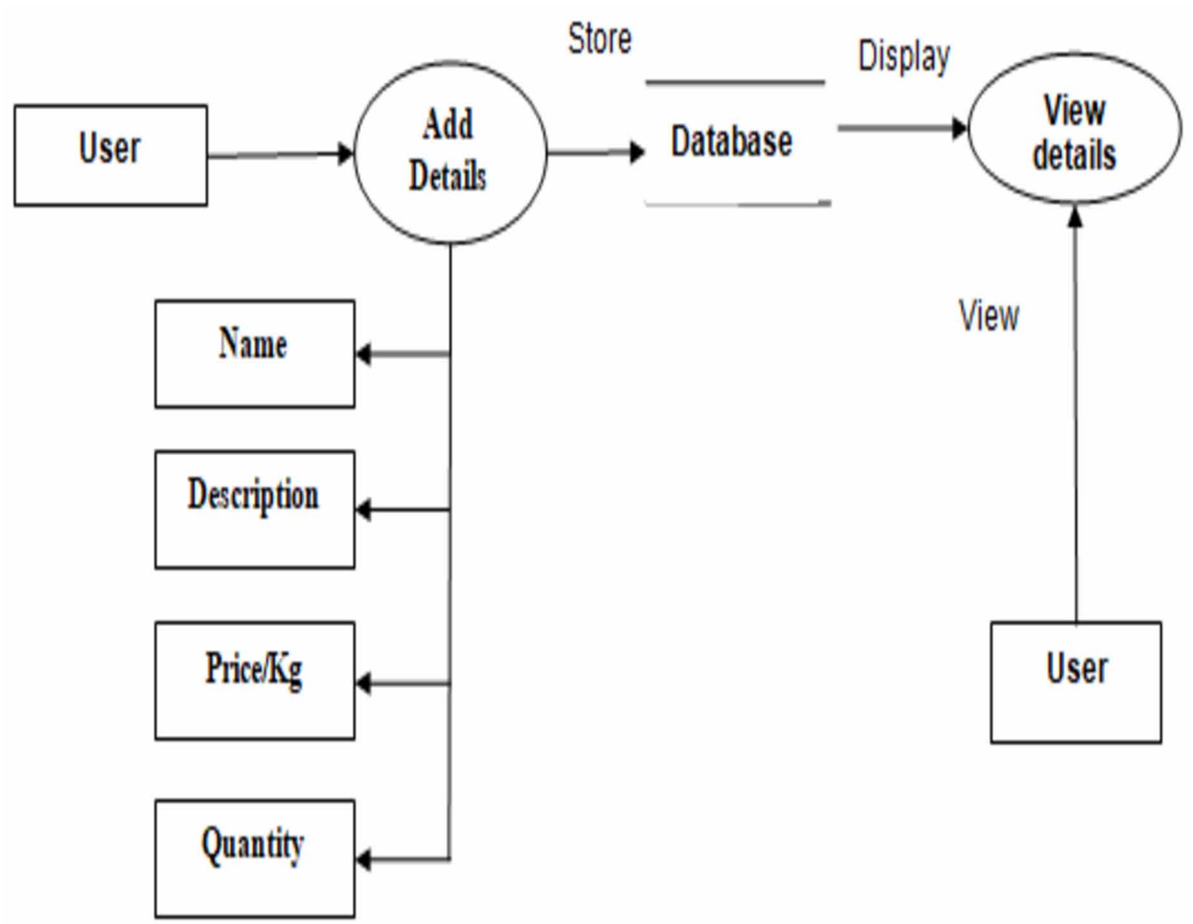


Apple images or picture data is captured against 34 variety apples. Percentage rate of colors red and pale yellow and counts of all the dot appears apple skins is overall or average rate taken against four apples picture or image is captured against irregular angles. During, the ripeness levels are examined by average rate from three different or random mark on the apple. If the sweetness or ripeness level of apple is higher than 14 Brix, then it is accepted to be ripe or fit for the consumption. Or else, apple upon the sweetness level or ripeness level lesser than of 14 Brix is accepted to be not ripe.

Banana images or picture data is captured against 42 variety bananas. Percentage rate colors yellow, brown and green and counts of all dots appears on banana skins is overall or average value taken with two other banana picture or image captured against irregular angles. During, the ripeness levels are given by overall rate of three different marks on banana. If the sweetness or ripeness level of banana is higher than 17 Brix, then it is accepted to be ripe or fit for the consumption. Or else, apple upon the sweetness level or ripeness level lesser than of 14 Brix is accepted to be not ripe.

Melon images or picture data is captured with 6 variety melons. Percentage rate color green and yellow and counts of dots taken from melon skins is the overall value taken from four different melons image captured against and irregular angles. During, the sweetness or ripeness level is given by the

Figure 10. Data flow for sell product



average rate of three different random marks of melon. If the sweetness or ripeness level of melon is higher than 12 Brix, then it is accepted to be ripe or fit for the consumption. Or else, melons upon the sweetness level or ripeness level lesser than 14 Brix is assumed to be not ripe. Further the activities of the ripe track are described in the Figure 13.

CONCLUSION AND FUTURE SCOPE

Number of reasons has been evicted as the major reason behind the farmer's apathy, specifying like scarcity, dues and marketing. However intermediate complication is major difficulty at the present days, in order eliminate the concept of middlemen this application is constructed. Marketing of agriculture can be made effective and efficient if it is looked from the collective and integrative efforts from various quarters by addressing to farmers, middlemen, researchers and administrators. As the problems faced by farmers are increasing every day i.e. lack financial support, less jobs and improper equipment and machineries due to this the agricultural sector in the country is going down drastically. Hence to

Figure 11. Data flow for purchase product from the farmer

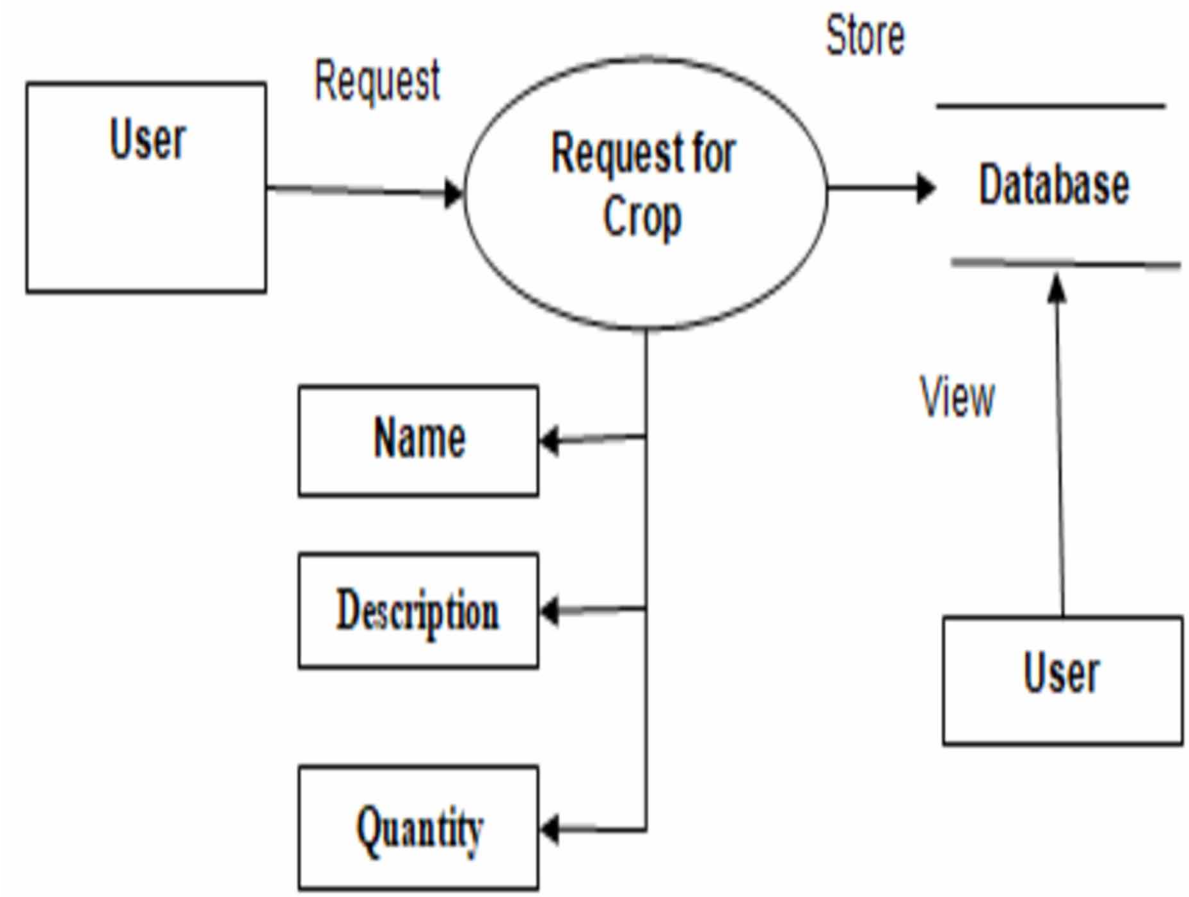


Figure 12. Fruit Images Captured from Various Angles (a) Apple (b) Banana (c) Melon

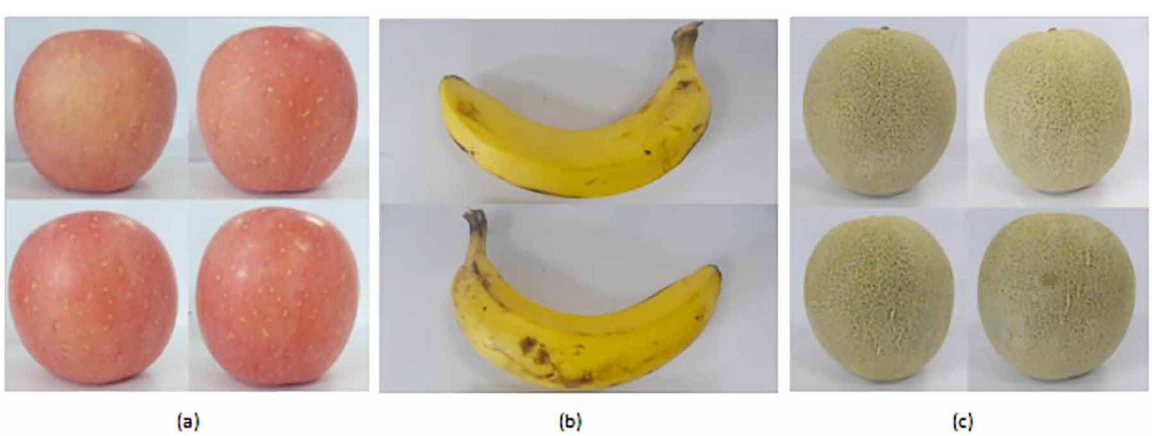
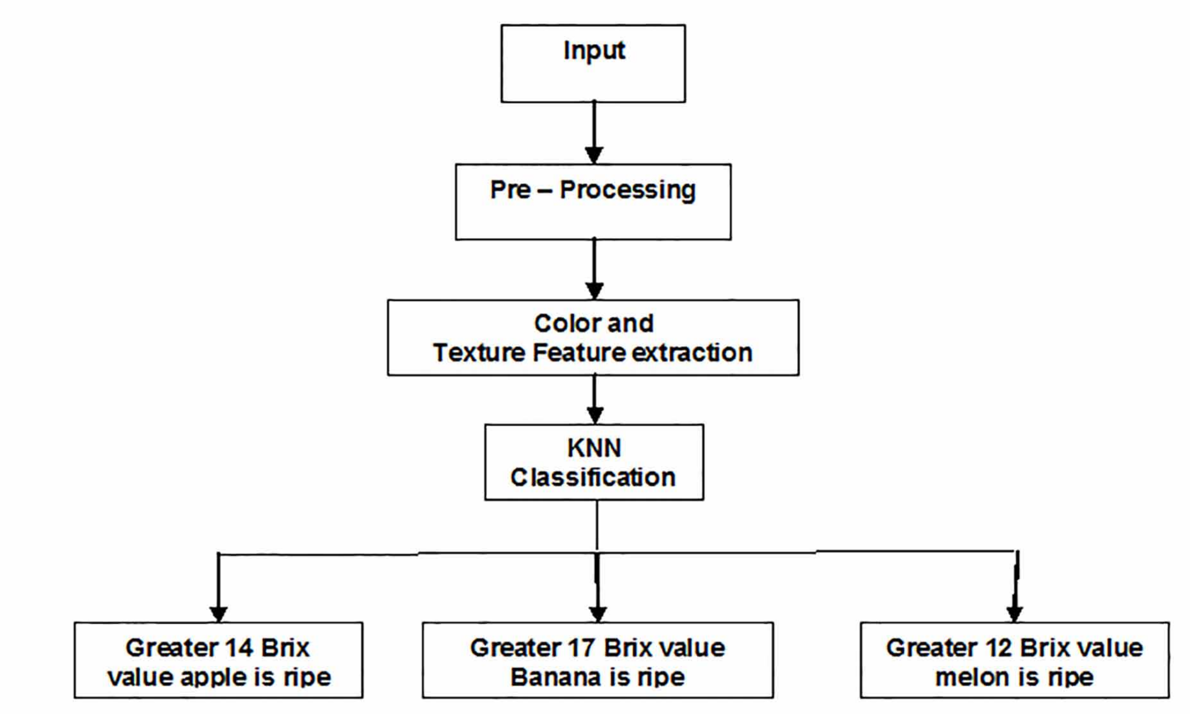


Figure 13. Activities for grouping the ripe track



overcome all these problems Thus the proposed system, an Android application that helps the farmers find suitable agricultural jobs, get the required machineries for rent at affordable price, obtain land on lease and receive the required financial support in the form of investment. This will help in the growth of agriculture sector in the country.

The future scope of the system is it can integrate different regional languages so that we can make the application multilingual. It can also provide services like decision support and expert advice. Feature like location supported search can also be implemented. Online selling of agro-based products like fertilizers and pesticides can be introduced. It can also display weather report and also forecasting of weather.

REFERENCES

Aggarwal, M., Kaushik, A., Sengar, A., Gangwar, A., Singh, A. K., & Raj, V. (2014). Agro App: An application for healthy living. *2014 International Conference on Information Systems and Computer Networks (ISCON)*, 30-32 10.1109/ICISCON.2014.6965213

Angadi, S. A., & Hatture, S. M. (2018). Hand Geometry Based User Identification Using Minimal Edge Connected Hand Image Graph. *IET Computer Vision*, *12*(5), 744–752. doi:10.1049/iet-cvi.2017.0053

Angadi & Hatture. (2011). A novel spectral graph theoretic approach to user identification using hand geometry. *International Journal of Machine Intelligence, 3*(4), 282-288.

Angadi & Hatture. (2019). Face Recognition Through Symbolic Modeling of Face Graphs and Texture. *International Journal of Pattern Recognition and Artificial Intelligence*. doi:10.1142/S0218001419560081

Chandra. (2011). Role of E-Agriculture in Rural Development in Indian Context. *International Conference on Emerging Trends in Networks and Computer Communications (ETNCC)*.

Forkuor, G., Landmann, T., Conrad, C., & Dech, S. W. (2012). Agricultural land use mapping in the sudanian savanna of West Africa: Current status and future possibilities. *2012 IEEE International Geoscience and Remote Sensing Symposium*, 6293-6296. 10.1109/IGARSS.2012.6352698

Gayatri, M., Jayasakthi, J., & Mala, G.S. (2015). Providing Smart Agricultural solutions to farmers for better yielding using IoT. *2015 IEEE Technological Innovation in ICT for Agriculture and Rural Development (TIAR)*, 40-43.

Ghadiyali, T., Lad, D. K., & Patel, B. (2011). Agriculture Intelligence: An Emerging Technology for Farmer Community. *2011 Second International Conference on Emerging Applications of Information Technology*, 313-316. 10.1109/EAIT.2011.36

Guptaa, S., & Trivedi, G. (2016). e-krishakMitra. *International Conference on Accessibility to Digital World (ICADW)*.

Iswari, N. M., & Wella, R. (2017). Fruitylicious: Mobile application for fruit ripeness determination based on fruit image. *2017 10th International Conference on Human System Interactions (HSI)*, 183-187.

Jayarathna, H. M., & Hettige, B. (2013). AgriCom: A communication platform for agriculture sector. *2013 IEEE 8th International Conference on Industrial and Information Systems*, 439-444.

Junfitrhana, A. P., Langlangbuana, M. L., & Fatah, W. A., & Susilawati (2017). Developing potential agriculture land detector for determine suitable plant using Raspberry-Pi. *2017 International Conference on Computing, Engineering, and Design (ICCED)*, 1-4. 10.1109/CED.2017.8308116

Karchi & Nagesh. (2017). A Hybrid Approach for Hyper Spectral Image Segmentation Using SMLR and PSO Optimization. *Proceedings of Recent Trends in Image Processing and Pattern Recognition, Springer Nature Singapore, CCIS 709*, 1–9.

Karchi, R. P., & Munusamy, N. (2018). Exploration of Unmixing and Classification of Hyperspectral Imagery. *International Journal of Future Generation Communication and Networking*, *11*(6).

Kassim, M. R., Mat, I., & Harun, A. N. (2014). Wireless Sensor Network in precision agriculture application. *2014 International Conference on Computer, Information and Telecommunication Systems (CITS)*, 1-5.

Khodaskar, K. (2015). Virtual Fruits Market – An Application for Farmer. *Fifth International Conference on Communication Systems and Network Technology*. 10.1109/CSNT.2015.235

Kumar, V., Dave, V., Nagrani, R., Chaudhary, S., & Bhise, M. (2013). Crop cultivation information system on mobile devices. *2013 IEEE Global Humanitarian Technology Conference: South Asia Satellite (GHTC-SAS)*, 196-202. 10.1109/GHTC-SAS.2013.6629915

Mhaski, R. R., Chopade, P. P., & Dale, M. P. (2015). Determination of ripeness and grading of tomato using image analysis on Raspberry Pi. *2015 Communication Control and Intelligent Systems*, 214–220.

Ministry of Agriculture & Farmers Welfare. (2015). *Kisan Suvidha.* Author.

Mohamad, M. R. A., & Gombe, M. I. (2017). *E-agriculture Revisited: A Systematic Literature Review of Theories, Concept,Practices, Methods, and Future Trends*. British Academy of Management.

Mondal, A. S., Misra, I. S., & Bose, S. (2017). Building a low cost solution using wireless sensor network for agriculture application. *2017 International Conference on Innovations in Electronics, Signal Processing and Communication (IESC)*, 61-65. 10.1109/IESPC.2017.8071865

Murakami, Y. (2014). iFarm: Development of Web-Based System of Cultivation and Cost Management for Agriculture. *2014 Eighth International Conference on Complex, Intelligent and Software Intensive Systems*, 624-627. 10.1109/CISIS.2014.89

Mutiawani, V., Subianto, M., & Makruf, M. (2017). An Android agricultural commodity price information application by utilizing RESTful web service. *2017 International Conference on Electrical Engineering and Informatics (ICELTICs)*, 109-113. 10.1109/ICELTICS.2017.8253249

Rajeswari, S., Suthendran, K., & Rajakumar, K. (2017). A smart agricultural model by integrating IoT, mobile and cloud-based big data analytics. *2017 International Conference on Intelligent Computing and Control (I2C2)*, 1-5. 10.1109/I2C2.2017.8321902

Rashmi & Munusamy. (2018). Hyperspectral Image Classification and Unmixing by using ART and SUnSPI Techniques. *International Journal of Database Theory and Application, 11*(3), 13-28.

Rashmi & Nagesh. (2013). A Review of Spectral Unmixing Algorithms in the context of Mars Dataset. *International Journal of Latest Trends in Engineering and Technology*, 55-60.

Sawant, A., & Durbha. (2014). Krishisense: A Semantically Aware Web Enabled Wireless Sensornetwork System for Precision Agriculture Applications. *IEEE Geoscience and Remote Sensing Symposium*.

Singhal, M., Verma, K., & Shukla, A. (2011). Krishi Ville — Android based solution for Indian agriculture. *2011 Fifth IEEE International Conference on Advanced Telecommunication Systems and Networks (ANTS)*, 1-5. 10.1109/ANTS.2011.6163685

Thankachan, S., & Kirubakaran, D. (2014). A Survey Conducted on E-Agriculture with Indian Farmers. *International Journal of Computer Science and Mobile Computing*.

Chapter 9

A Study on Technology-LED Solutions for Fruit Grading to Address Post-Harvest Handling Issues of Horticultural Crops

Arun Kumar R.
KLS Gogte Institute of Technology, India

Vijay S. Rajpurohit
KLS Gogte Institute of Technology, India

Sandeep Kautish
LBEF Campus, Nepal

ABSTRACT

The reduction of post-harvest losses and value addition of the horticultural corps has attained the higher priority of the current research works. Grading is the major phase in post-harvest handling. Presently grading is done on the basis of observation and through experience. Various drawbacks associated with such manual grading are subjectivity, tediousness, labor requirements, availability, inconsistency, etc. Such problems can be alleviated by incorporating automation in the process. Researchers round the clock are working towards the development of technology-driven solutions in order to grade/sort/classify various agricultural and horticultural produce. With the motto of helping the researchers in the field of grading and quality assessment of fruits and other horticulture products, the present work endeavors the following major contributions: (1) a precise and comprehensive review on technology-driven solutions for grading/sorting/classification of fruits, (2) major research gaps addressed by the researchers, and (3) research gaps to be addressed.

DOI: 10.4018/978-1-5225-9632-5.ch009

1. INTRODUCTION

Horticulture, the branch of agriculture, can be defined as the science and art of cultivating and marketing fruits, vegetables, flowers, nuts and ornamental plants. The word horticulture can be split into two Latin words viz. *hortus*, meaning 'garden' and *cultus*, meaning 'tilling'. Horticulture can be distinguished from that of agriculture by its scale of production and commercialization. Horticulturists usually apply their skills and technologies to cultivate intensively produced plants for human food purposes and allied social needs. Horticultural produce imparts a vital part in human nutrition as they supply the necessary vitamins, minerals, dietary fiber and anti-oxidants. Horticulture sector has become one of the major drivers of the growth in the agriculture sector (Horticulture Statistics Division, 2017).

At its widest scale, post-harvest horticulture involves transformation of the product from the state it is disjointed from the plant or growing medium to the state of ready-to-consume by the end consumer. The field of post-harvest includes various stages / states / processes such as cleaning, sorting, grading, removal of field-heat, waxing, packing, storage and transportation. Irrespective of the scale or complexity, post-harvest activities add 'value' to the product (Collins, R. J., 2009).

Among all the post-harvest handling processes, sorting and grading plays a vital role. This is because of the reason that the sorting and / or grading is the stage that actually involves endeavoring the true market value of the commodity.

The quality and safety of the horticulture products that reach the end consumer hinges due to poor practices of post-harvest management. Both qualitative and quantitative losses ensue in horticultural commodities in between the harvest and the consumption. The quantitative losses are easy to assess. But the qualitative losses are difficult to assess as they include the aspects such as caloric value, edibility, consumer acceptability, nutritional quality etc. Moreover, the quality standards and consumer preferences differ prominently across countries and across cultures. These differences in turn affect marketability and the extent of post-harvest losses. Post-harvest losses fluctuate greatly depending on the type of commodity, production areas and the production season.

Reduction of the post-harvest losses has many benefits. Few of them are as under (Department of Horticulture, 2014):

1. It can surge the food availability to the growing population of the world.
2. Improves food security and imparts better nutrition to the consumers.
3. Reduction of the losses reduces the area needed for production.
4. Reducing post-harvest losses also conserves natural resources.
5. Boosts up the financial stability of the growers / farmers.
6. Boosts up the financial stability of the exporters / traders / processors/ transporters.
7. Helps to uplift the financial growth of the country.

The present article is organized as follows. Section 2 discusses about fruits and their importance in general. Various issues of post-harvest handling with respect to grading of fruits are discussed in section 3. Section 4 throws light on the importance of incorporating technology in addressing the post-harvest issues as mentioned in section 3. Section 5 discusses in detail about the available literature in connection to solving the post-harvest handling issues with respect to grading of fruit with technology driven solutions. Section 6 discusses the major issues being addressed and section 7 discusses the issues and research gaps to be addressed. Finally, the chapter is concluded in section 8.

2. FRUITS AND THEIR IMPORTANCE

Fruits play a significant role in human nutrition as they are the essential sources of vitamins, minerals, dietary fiber, and antioxidants. Consumption of a diverse fruits on daily basis is highly endorsed because of the associated health benefits. Fruits are the rich sources of ascorbic acid having beneficial effects of wound healing and antioxidant. Dietetic source of vitamin C is essential as human beings can't synthesize it. Some of the fruits are rich in beta carotene, which is an essential supplement to maintain good sight health. Fruits avoid people suffering from degenerative diseases. Phenolic compounds, antioxidants and nutritive fiber are the essential elements in sinking the risk of several types of cancers. Fruits those are rich in potassium help to maintain the blood pressure. Fiber from fruits helps in reducing blood cholesterol levels and hence aids in lower risk of heart diseases. Fiber also helps in reducing constipation and diverticulosis. Fruits those are rich in fiber reduces the risk of obesity and type 2 diabetes. Due to their high nutritive value, ready availability and being inexpensive they significantly contribute to human well-being.

3. ISSUES IN POST-HARVEST GRADING OF FRUITS

As mentioned earlier, among all the post-harvest handling processes, sorting and grading plays a vital role in devising the true market value of the fruits. At present grading of many fruits is carried out on the basis of observation and experience. Poor judgment ensues due to the following various reasons (Xiaobo, Z. et al. (2007), Omid, M. et al. (2010) and Mustafa, N. B. A. (2009)):

1. Seasonal fluctuations in the grading criteria.
2. Differences among production areas.
3. Intensive labor requirements during the post-harvest processes.
4. Fatigue of grading workers.
5. Subjective nature of the grading process.
6. Tediousness of the workers throughout the process.
7. Availability of the labors.
8. Higher costs / wages.
9. Inconsistency due to the presence of human in the loop

Post-harvest of the fruits experience huge losses due to the above-mentioned reasons. It has been estimated that about one third of the horticultural crops produced are never consumed by humans (Kader, A., and Rolle, R., 2004). In developing countries like India, 20% to 30% of the horticultural produce is lost beforehand consumption due to poor harvesting, management, storage, transportation and marketing practices. The major goal of the horticultural division of Indian Council of Agricultural Research (ICAR) is to inculcate the technology-led solutions in the various processes of horticulture. It has directed to monitor Research & Development programs of National interest in supporting the knowledge repository of the horticulture division.

4. IMPORTANCE OF FRUITS GRADING USING TECHNOLOGY IN PLACE

Post-harvest losses of fruits are difficult to predict. It has to develop in relation with needs of each society to inspire production, prevent post-harvest losses, improve nutritional quality and add value to the production. Few of the strategies for loss prevention may include the following:

1. Usage of genotypes that abide lengthier post-harvest-life.
2. Usage of integrated crop management systems resulting into good quality of the products.
3. Incorporating good practices in the agricultural production resulting into good retention of the quality of the fruits.
4. Inculcating Information and Communication Technology (ICT) throughout various practices of post-harvest handling of the fruits; especially during sorting / grading of the commodity.

It is very much necessary to grade the agricultural produce brought to the market. Advances in post-harvest technologies and ICT in place, it is possible to create new products, new processes and newer ways of managing information. The technology has advanced from year to year where the manual or labor-intensive works undertaken and managed by humans is substituted by the automated systems accomplished by machines. Automated grading and quality assessment of agricultural produce has attained an exceptional attention due to ever increasing demand for high quality products to be produced in a shorter time period (Jamil et al., 2009). Automatic sorting of fruits with the help of machine intelligence and soft computing techniques can improve the quality of the product, eliminate the unreliable manual evaluation; there by reducing the enslavement of the manpower required in the process (Iraji et al., 2011). The automated work basically results in increase of speed. The automation is not limited to industrial specific things. It has spread to the agriculture arena too. Machine vision-based sorting / grading of fruits is necessary to upsurge the speed of the process and also to exterminate any human error in the process (Unay & Gosselin, 2005).

Market grade of quality food products are determined on the basis of their multiple features such as appearance, texture and flavor. Flavor can be measured on the basis of chemical methods; but it is a destructive one. Properties such as mouth feel are tough to measure. Machine vision is the most important tool and technique for measuring the external features such as size, shape, color, bruises etc. (Omid et al., 2010). The external appearance is the fundamental factor in pricing any fruit. Grading of fruits plays a vital role in the recent days. An attempt can be made to perform such a grading through non-destructive methods by employing machine vision in place.

In the following section, a detailed study has been outlined that discusses the previous works carried out by various researchers in adopting machine vision for the purpose of grading and quality assessment of various fruits.

5. TECHNOLOGY-DRIVEN SOLUTIONS IN GRADING FRUITS

In this section a detailed study on the existing machine vision-based solutions for the purpose of grading various fruits is outlined. The study comprises of the survey carried out for various fruits for the purpose of grading and quality assessment.

5.1. Apple

Xiaobo, Z. et al. (2007) proposed a novel Organization Feature Parameters (OFP) based method for grading apples into four classes viz. (1) Class 'Extra', (2) Class 'Class I', (3) Class 'Class II' and (4) Class 'Reject'. A total of 318 apple images were captured using a custom-built machine vision system as depicted in Figure 1.

About 95% surface area of the fruit was obtained by capturing the fruit from different angles for each fruit. 200 samples were used as the 'Training set' and the remaining 118 samples were used as the 'Testing set'. 17 color feature parameters were extracted for each image and were normalized. A step decision tree was used for assessing the grades of apples using Organization Feature Parameter (OFP) with genetic program. The results were obtained using two of the machine learning techniques viz. Back Propagation Artificial Neural Networks (BP-ANN) and Support Vector Machines (SVM). Results of the OFP method were found to be more accurate than BP-ANN and quite lower than SVM.

Unay, D., and Gosselin, B. (2005) developed a computer vision based system to automatically grade apple fruits. Database consisted of one-view images of 'Jonagold' apples. Images were captured from diffusely illuminated environment using a high resolution monochrome digital camera. The image acquisition was done with four interference band-pass filters that were centered at 450nm, 500nm, 750nm, and 800nm having the bandwidths of 80nm, 40nm, 80nm, and 50nm respectively. Figure 2 shows some of the sample defected apples.

Figure 1. Schematic view of the machine vision system
(Courtesy: Xiaobo, Z. et al., 2007)

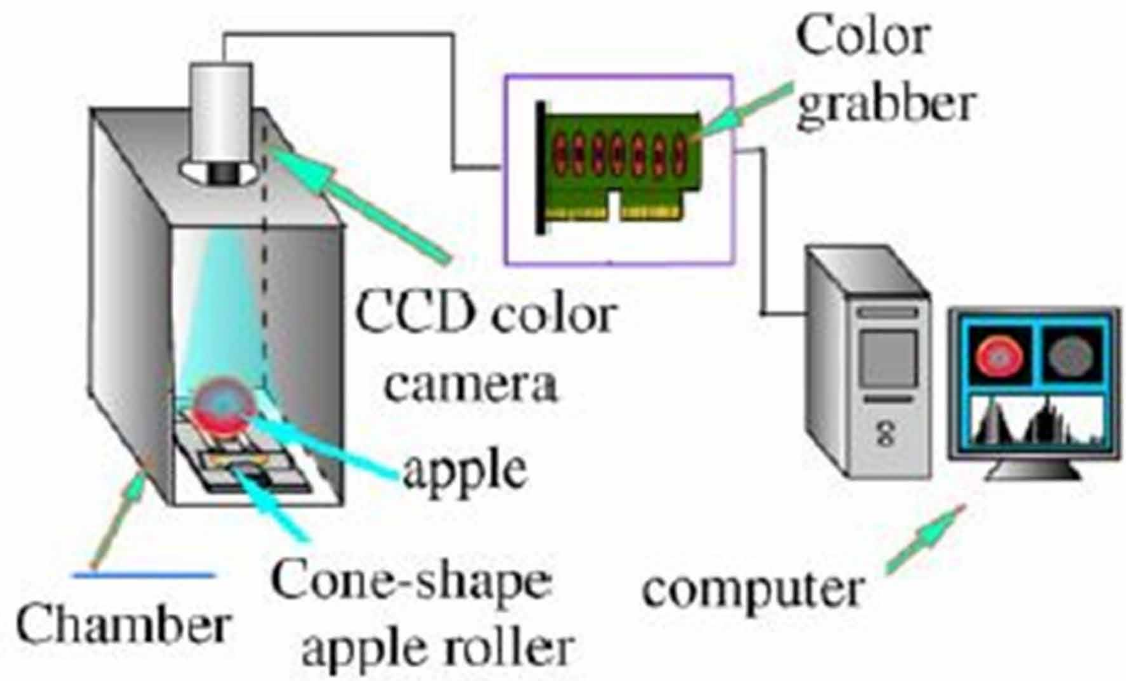
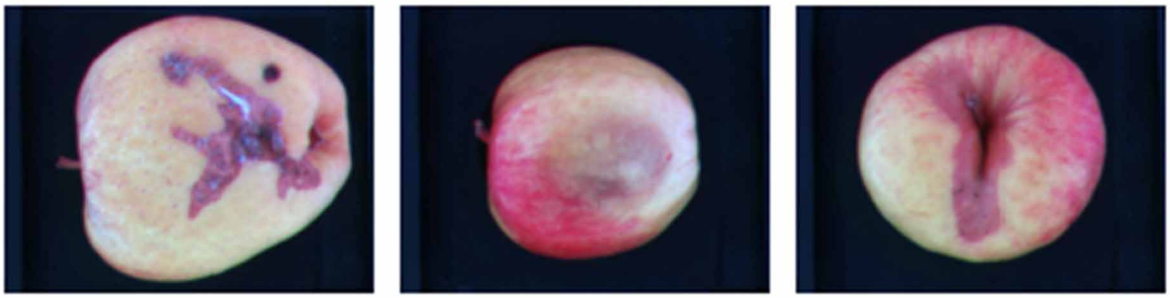


Figure 2. Some defected apples
(Courtesy: Unay, D., and Gosselin, B., 2005)



Segmentation of defected skin was performed by 3 techniques of global thresholding viz. (1) Otsu (2) Isodata and (3) Entropy. Stem end or calyx regions were eliminated as they were incorrectly classified as defects. Statistical features were then extracted from segmented defects. The features were then fed into the following supervised classifiers: (1) Linear Discriminant Classifier (LDC) (2) K-Nearest Neighbor Classifier (k-NN) (3) Fuzzy k-Nearest Neighbor Classifier (fuzzy K-NN) (4) Adaptive Boosting (AdaBoost) and (4) Support Vector Machines (SVM). The highest recognition rate was approved by support vector machine classifier that demonstrated an accuracy of 89.2%.

Unay, D., and Gosselin, B. developed a machine vision system to grade Jonagold apples (2005). Images of the apples were obtained under diffuse light source at different wavelengths. The images acquired consisted of both healthy and skin defected. A total of 280 healthy fruits and 246 skin defected fruits were imaged. The acquired images were then fed into a pre-processing step where in thresholding and morphological filling operations were applied. Back Propagation Artificial Neural Networks (BP-ANN) were employed for defect segmentation. A total of thirteen statistical features were extracted for each of the fruit image. Five supervised classifiers were employed to discriminate the apples. They were: (1) Linear Discriminant Classifier (LDC) (2) K-NN (3) Fuzzy K-NN (4) Adaptive Boosting and (5) SVM. Results of the experiments demonstrated a high accuracy of 90.3% with SVM.

Apple fruit grading was carried out with the help of digital image processing by Dang, H. et al. (2010). The apple fruits were imaged using a CMOS camera. Images were then fed into a pre-processing module where in a series of pre-processing steps were applied. Images were first filtered using a median filter. Gray scale conversion was then performed. Edge detection of the fruits was carried out using OTSU's method. Size of the fruits was then detected based on the symmetry of the fruit. Finally grading of the fruit was done based on the size of the fruit. Results of the experiments were satisfactory with less than 4% errors along with a processing speed of about 1.5 s.

Maturity level of apple fruits were graded into five classes by Dadwal, M., and Banga, V. (2012). The RGB color space features were used as the criteria. Color image segmentation was performed for the acquired RGB images. The mean values of the Red, Green and Blue channels were then extracted. Fuzzy logic was then employed in order to determine the ripeness class of the fruits on the basis of the extracted color features. The results obtained were satisfactory.

Leemans, V. et al. (2002) performed grading of apples according to the European standard. Two varieties of apples were considered, they were golden delicious and jonagold. These apples were graded into four quality grades, they were: (1) Extra class (2) Class I (3) Class II and (4) Reject class. Image acquisition of the fruits was done with two cameras. Authors employed groups of parameters to characterize the defects of the apples. They were: geometrical parameters, color parameters, texture parameters and calyx & stem characterizing parameters. Grading was carried out using two methods and the results were compared. Quadratic Discriminant Analysis and Artificial Neural Networks were employed. The average results of the experiments were 78% for golden delicious and 72% for jonagold apples.

Apples were graded into five quality categories by Lorestani, A. N. et al. (2006). The apple fruits were imaged using an acquisition unit as depicted in Figure 3.

The RGB color features and the size features were extracted and were fed into a fuzzy inference system for the purpose of classification. The results of the experiments showed an accuracy of 90.8% as compared to the classification done by human expert.

Classification of the golden delicious apples according to USDA standards was carried out by Kavdır, I., and Guyer, D. E. (2008). Nine features were extracted for each fruit. Four of the classification algorithms were employed and results were compared. The classifiers employed were: (1) Plug-in Decision Rule (2) K-Nearest Neighbor (3) Decision Tree classifier and (4) Artificial Neural Networks. Classification was done by using 4 features, 5 features and 9 features. The highest classification success rate was achieved by neural networks that demonstrated an accuracy of 90%.

Jonagold apples were graded real time by Leemans, V., and Destain, M. F. (2004). Altogether 16 features were extracted that characterize the defects of the apples. Color, shape, texture and position features were extracted. Quadratic discriminant analysis was employed for grading the apples which demonstrated the correct classification rate of 73%.

Figure 3. Acquisition system used for apple sorting
(Courtesy: Lorestani, A. N. et al., 2006)

Nakano, K. (1997) developed an application of artificial neural networks to grade apples in order to reduce the disadvantages associated with manual grading. The grading was based on the color parameters of the apple images. Images were acquired using a CCD video camera. There were two ANN models being adopted, one being used to classify the pixel and another being used to grade the whole apple. Whole apple was graded into one of the following classes: (1) Superior (2) Excellent (3) Good (4) Poor colored and (5) Injured. Results of the experiments were satisfactory with the judgment ratio of 95%.

5.2. Banana

Mustafa, N. A. et al. (2008) developed a machine vision system to determine the size and ripeness of banana. Images of a bunch of banana were captured. The captured images were first gray converted. Exterior outline of the banana bunch was then determined by applying Canny's edge detection method. The region properties were then extracted that included area, perimeter, length and thickness. Color based segmentation with L*a*b* color space conversion was used for the calculation of the ripeness percentage of the fruits along with a reference object. Figure 4 depicts a sample banana fruit with the corresponding ripeness percentage.

5.3. Cashew

Narendra, V. G., and Hareesha, K. S. (2011) developed a machine vision system to classify cashew kernels into six classes. Images of cashew kernels were obtained under the artificial lighting conditions using a digital camera. Image segmentation was then performed with thresholding technique. Color features were extracted for each image and fed as input to the back propagation artificial neural networks. Results of testing showed an overall accuracy of 80%.

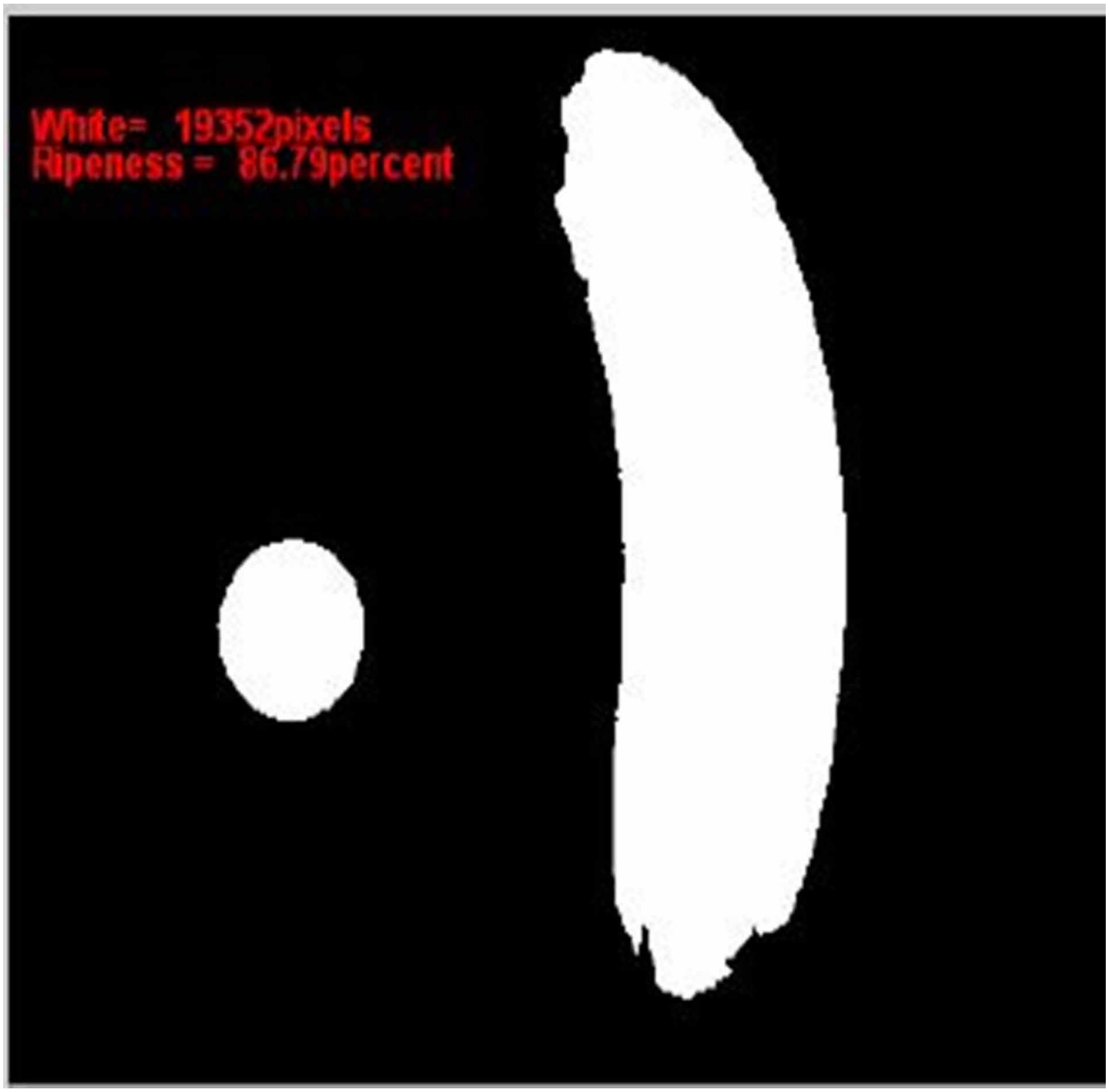
5.4. Dates

Lee, D. J. et al. (2008) proposed a unique methodology for evaluating quality of the date fruits. Two techniques were employed for quality evaluation of the Medjool dates viz. Color quantization and color analysis. Authors did not consider either RGB or HIS due to their few of the drawbacks associated with the color representation and analysis. An image dependent color quantization process was carried out with calibrations in order to find the coefficients for color analysis. The accuracy of the results were highly satisfactory with values of 92.5% in categorizing red dates, 82.8% in categorizing orange fruits and 88.7% in classifying surface defects. These results were with respect to the visual inspection by a human expert.

5.5. Jatropha

ANN based pattern recognition system was developed by Effendi, Z. et al. (2009) in order to determine the maturity grade of the Jatrophacurcas fruits. The fruit type, fruit size, skin color and size of defects were used in order to grade the fruits. Three quality categories were considered viz. (1) Raw (2) Ripe and (3) Over ripe. Back propagation algorithm of artificial neural networks was employed for the purpose of training the features extracted out of the fruit image. The results of the experiments were satisfactory.

Figure 4. Ripeness percentage of a banana fruit
(Courtesy: Mustafa, N. A. et al., 2008)

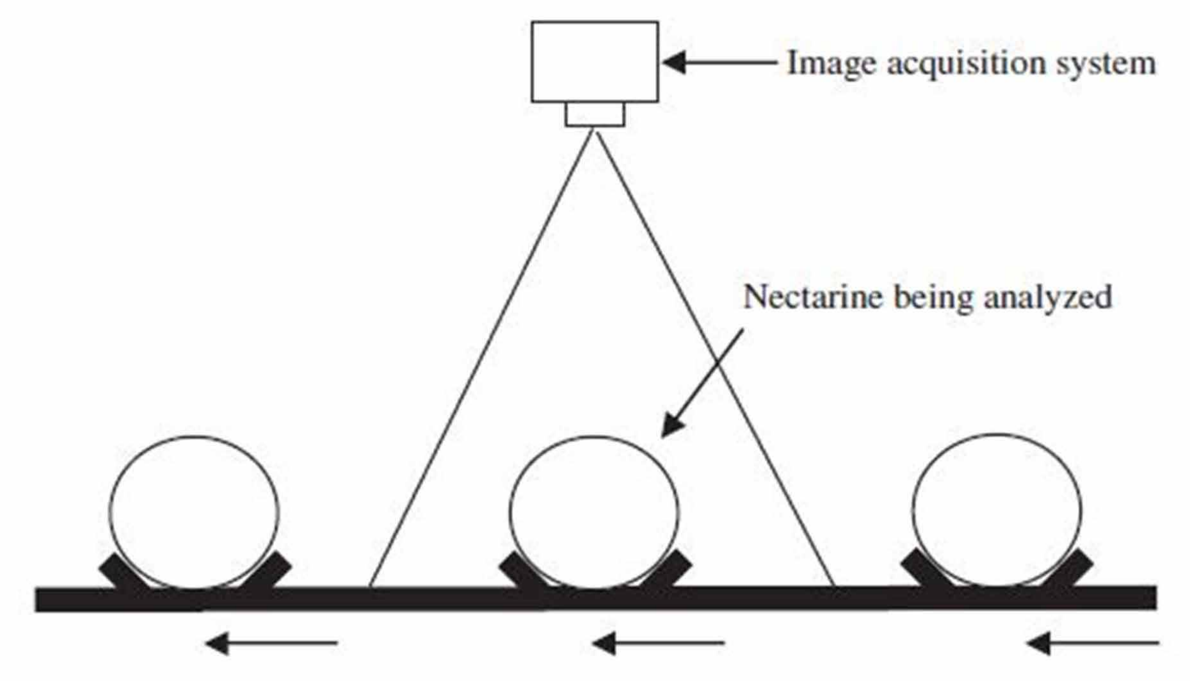


5.6. Nectarine

Aim of the research work carried out by Font Calafell, D. et al. (2014) was to design and develop a machine vision system for the verification of individual nectarine varieties. The system was intended to be used in the industrial fruit packing lines as depicted in the Figure 5.

The application was developed to verify Necterreve nectarine among all other varieties viz. ASF 08.05, Honey BlazePVC, ASF 08.21, Queen Gem, Nectarreve PVCR (Plant Variety Certificate Requested), NectareinePVC (Plant Variety Certificate). The method started with an image acquisition unit operated

Figure 5. Schematic view of the proposed packing line for Nectarines
(Courtesy: Font Calafell, D. et al., 2014)



with constant diffuse illumination. The non-interpolated RGB color images, within an ROI of 240-by-240 pixels, were obtained from the acquisition system. The algorithm was carried out in the following four steps: (1) Segmentation based on Otsu method (2) Gray conversion (3) Defining an inner circular ROI with a radius of 80 pixels at the center of the nectarine and (4) Computing and analysis of feature histogram vector. Two comparison methods were used viz. (1) Pearson's correlation coefficient and (2) Cumulative absolute subtraction (or Manhattan distance). Results of the proposed system showed an accuracy of 87% compared with that of a human operator.

5.7. Oil Palm

Nursuriati Jamil et al. (2009) developed an automated grading system to grade palm oil Fresh Fruit Bunches (FFB) in order to improve accuracy and quality grading of FFB in palm oil mills. The FFB were classified into three grades viz. (1) Over ripe (2) Ripe and (3) Unripe. Two grading methods were applied and the results were compared. The methods applied were: (1) Color grading using RGB digital numbers (DN) and (2) Color grading trained using a supervised learning Hebb technique and classification using fuzzy logic. A total of 90 FFB images were captured under the direct sunlight with the help of a commercial digital camera. A user-created mask was used for the removal of background. Hebb algorithm was implemented to identify the best-fit value that represents RGB color of the FFB images. Color classification was done in four steps viz. (1) Fuzzification (2) Rule evaluation (3) Aggregation of the rule outputs and (4) Defuzzification. 45 FFB images were used in the testing phase in order to compare the results of DN and Neuro-Fuzzy algorithm. Color grading with RGB Digital Numbers showed an accuracy of 60% in grading over ripe FFB, 27% for ripe FFB and 60% for unripe bunches. Color Grading using Neuro-Fuzzy System demonstrated an accuracy of 80% in grading over ripe FFB, 73.3% in grading ripe FFB and 66.7% for under ripe FFB. Analysis of the results illustrated that Neuro-Fuzzy System outperformed the DN method by achieving an accuracy of 73.3%.

Oil palm fruits were graded into three quality grades using a machine vision system by May, Z., and Amaran, M. (2011). A total of 75 images of oil palm fruits under were obtained under controlled laboratory conditions. Three classes of grades were considered viz. (1) Under ripe (2) Ripe and (3) Over ripe. The acquired images were scaled to 640*480 resolutions. Background subtraction method was applied in order to obtain the object of interest. Color features were extracted for each of the image using RGB color model in which mean of each channel in RGB was obtained for each image. Fuzzy Inference System (FIS) was built in order to grade the oil palm fruits. There were three inputs (range of each channel in RGB) and one output (one of the three grades). Trapezoidal membership functions were used and fuzzy 17 rules were defined. Experimental results were compared against and expert human operator which showed an accuracy of 86.67%.

5.8. Orange

Artificial Neural Networks (ANNs) were employed to perform the quality assessment of Iyokam orange fruits by Kondo, N. et al. (2000). Images of fruits were acquired using a color TV camera. Three classes of features were extracted viz. color features, shape features and roughness features. The extracted features were then fed as input to the ANNs to predict the sugar content and pH contents of the image. The accuracy was found to be 94% in predicting sugar content and 81.9% in predicting pH content. Therefore, the overall accuracy was 87.95%.

5.9. Papaya

Shape based classification of the papaya fruits was carried out by Riyadi, S. et al. (2008). Images of the papaya fruit were acquired using a camera. Three of the image pre-processing steps were applied viz. (1) Image resize (2) Image segmentation using Otsu's method and (3) Boundary detection. The pre-processed images were then transformed using Discrete Wavelet Transform (DWT). Statistical properties of the sub bands of the DWT transformed images were then extracted. Feature fusion technique was then adopted and Linear Discriminant Analysis (LDA) method was adopted for the purpose of classification. Results of the experiments demonstrated an accuracy of 98%.

5.10. Raisins

Omid, M. et al. (2010) designed and developed an efficient machine vision system for grading bulk raisins. Two grade classes were considered: (1) The Desirable class and (2) The Undesirable class. Images of the raisins were acquired using a video camera. Acquired images were first background subtracted. The image processing algorithm was developed to extract RGB color features, size feature and center of gravity. A confusion matrix was obtained by comparing the proposed algorithm against human expert. The results demonstrated a Correct Classification Rate (CCR) of 96%.

5.11. Rapeseed

Objectives of the work carried out by Kurtulmuş, F., and Ünal, H. (2015) were (1) To develop a computer vision based expert system to differentiate seven varieties of rapeseed and (2) To conclude with the best predictive model along with optimum feature set. Image acquisition was done using regular office scanner with mass surface imaging technique. A total of 525 sample images were obtained and 420 color texture features were extracted for each sample. The features extracted included 14 Gray Level Co-occurrence Matrix (GLCM) features, 11 Gray Level Run Length Matrix (GLRLM) features and 59 Local Binary Pattern (LBP) features for each image. The features were then normalized to a scale of 0 and 1. Three feature models were adopted in order to reduce the number of features for training. They were (1) Stepwise Discriminant Analysis (SDA) using Gram-Schmidt process, in which the feature set was reduced to 44. (2) Principal Component Analysis (PCA) and (3) Recursive Feature Elimination (RFE) in which the feature set was reduced to 29. Three of the machine learning techniques were used for the classification purpose. They were (1) Support Vector Machine (SVM) (2) K-Nearest Neighbor (K-NN) and (3) Stochastic Gradient Descent (SGD). Results were analyzed using the confusion matrix that showed an accuracy of 99.24% with RFE and SVM techniques.

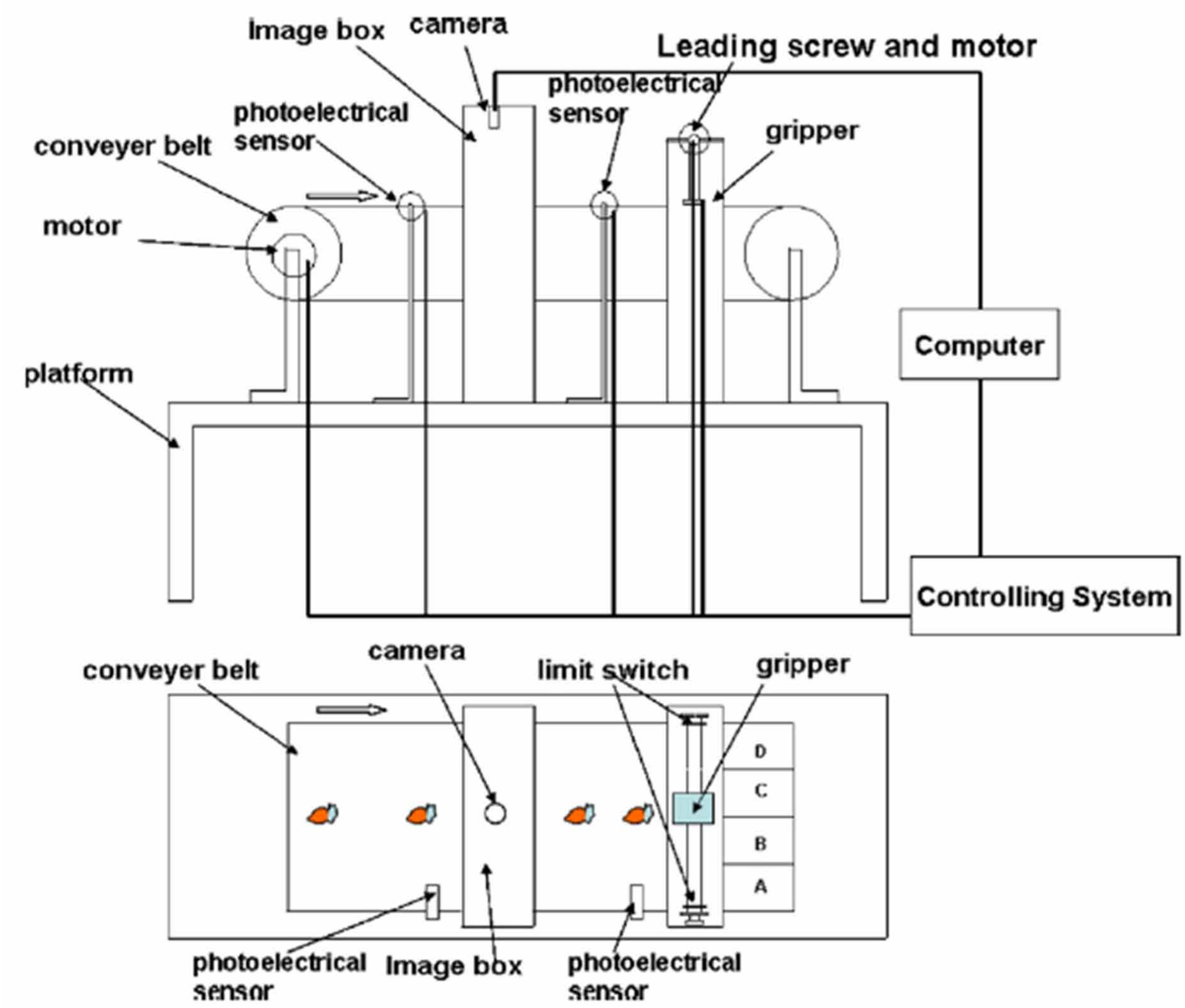
5.12. Strawberry

Liming, X., and Yanchao, Z. (2010) developed a machine vision system for the purpose of automating the grading process of strawberry fruits. A novel algorithm was designed developed to calculate the shape, size and color of the strawberry. This data was then used to grade the strawberry fruits. The structure of the strawberry grading system developed is depicted in Figure 6.

The system consisted of the following four units: (1) The mechanical unit (2) An image processing unit (3) The detection unit and (4) The control unit. The mechanical unit consisted of 1. A conveyer belt 2. Platform 3. Leading screw 4.Gripper and 5.Two motors. Motors were used to assist the transport of strawberry and hence the gradation process. The image processing unit consists of (1) Camera (Model: WV-CP470, Panasonic) (2) Image collection card (3) Shut image box and (4) Computer. Strawberry grading was carried out on the basis of three types of features viz. shape, size and color. Shape gradation was based on long-taper, square, taper and rotundity. Size gradation was implemented based on a threshold value corresponding to the strawberry maximum horizontal diameter. For color gradation, L*a*b* color space was used in which the threshold value is set according to the a* channel. The results obtained were satisfactory with a less than 5% size error, an accuracy of 88.8% with color grading and accuracy of 90% with shape grading. Also, the algorithm took a maximum ofthree seconds to process a fruit.

Aim of the study carried out by Yamamoto, K. et al. (2015) was of twofold: (1) To develop an automated system to evaluate the quality of the strawberry fruits and (2) Cultivar identification of strawberry fruits. Quality evaluation of the fruits was based on three of the appearance characteristics viz. color, shape and size. A total of 34 cultivars were identified using a machine vision-based method. Initially, images of the fruits were acquired using a chromatic image capturing system. All the acquired images were converted from RGB space to HSV color space. Color analysis was carried out using a Modified Improved - Color Distribution Entropy (I-CDE) method. The I-CDE methods was suitable to analyze images of one object i.e., strawberry. Shape analysis was carried out using the shape analysis tool and

Figure 6. Structure of automated strawberry grading system
(Courtesy: Liming, X., and Yanchao, Z., 2010)



Earth Mover's Distance (EMD) as the metric. Size analysis was based on the number of pixels in the image. Two statistical methods were used for the quality evaluation. They were: (1) Hierarchical Cluster Analysis (HCA) and (2) Multidimensional Scaling (MDS). Cultivar identification was carried out using Principal Component Analysis (PCA), Analysis of Variance (ANOVA) and Linear Discriminant Analysis (LDA) with leave-one-out cross validation method. Results of the study were < 42% when only single feature was considered and 68% after considering the combined features. The major contributions of the work were of twofold viz. (1) The system can be used for phenotyping purpose and breeding; because of the reason that it can evaluate any small differences that is typically recognized by professionals only. (2) The system can also be used as a new index of the appearance characteristics.

5.13. Tomato

Iraji, M. S. et al. (2011) proposed a method for sorting of tomatoes in order to reduce the dependency on the available manpower. Four categories of tomatoes were considered viz. (1) Export quality (2) National quality (3) Regional and (4) Salsa quality. Images of tomato were captured using a camera. The images were then made to undergo pre-processing steps that includes reflection reduction, contrast improvement and features extraction. The extracted features were used for analyzing the defects. Two methods were employed for classification viz. (1) Fuzzy Mamdani inference and (2) Adaptive Fuzzy Neural Network (ANFIS). Results demonstrated more accuracy with ANFIS.

Arefi, A. et al. (2011) developed a new segmentation algorithm to assist a robotic arm in picking ripen tomato. A Total of one hundred and ten color images of greenhouse-grown tomato variety were captured in the RGB color space. A combination of HSI and YIQ models was employed in order to extract the ripen tomato. Morphological operations were then applied in order to localize the extracted ripen tomato. A total of forty random images were used for the purpose of testing. The accuracy of the algorithm was found to be 96.36% with a processing time of 2.2 second per tomato image.

A novel tomato fruit recognition system was proposed by Yang, L. et al. (2007) in order to automatically harvest the fruits. The recognition system was built by adopting a Color Layer Growing segmentation technique followed by reconstruction. The 3-D reconstruction was carried out by using depth segmentation. Experimental results were satisfactory with respect to images having severe noise.

A low-cost machine vision based classifier was developed by Clement, J. et al. (2012) in order to classify tomato fruits. The weight of the fruit was used as the physical parameter. Image features extracted include diameter and color. Gradation of the fruits was carried out based on weight, diameter and color. Accuracy of the results were highly encouraging, and prototype of the system could perform 12.5 ratings per second.

Rokunuzzaman, M., and Jayasuriya, H. P. W. (2013) developed a digital image processing based low cost sorter for sorting tomatoes. Sorting was based on defect detection. Color features, shape features and the number of green objects in the image were used as three criteria for taking a sorting decision. Two of the techniques were adopted for the purpose of sorting viz. rule-based approach and Artificial Neural Networks (ANNs). Accuracy of the results demonstrated the values of 84% for rule-based approach and 87.5% for ANNs. The sorting of tomatoes was carried out with a high speed of 180 tomatoes per second.

5.14. Different Varieties of Fruits

NurBadariah Ahmad Mustafa et al. (2009) proposed a new technique of sorting and grading of fruits automatically. Four types of fruits and one variety of vegetable were used for conducting the experiments. They were 1. Apple 2. Banana 3.Orange 4.Mango and 5.Carrot. The methodology was built with the aim of overcoming the problems associated with manual grading. Three processes were employed viz. (1) Feature extraction (2) Sorting and (3) Grading. The feature values were then fed as input to Support Vector Machines (SVM). SVMs could sort the fruits based on their shape and size. Finally, fuzzy logic was employed to determine the grade of the fruit. The results were very promising for three of the five chosen fruits with an average accuracy of 95.78%.

Rocha, A. et al. (2010) introduced a machine vision system for multi-class fruits and vegetables classification for supermarkets. A total of 2633 images consisting of 15 classes of supermarket produce was collected and captured over a period of five months. The captured images were background subtracted with the help of a novel method based on K-Means algorithm. Digital image processing techniques were applied to extract the bag-of-features consisting of statistical color, texture and structural descriptors. The proposed method combined many features and classifiers. A novel feature and classifier fusion technique was employed. Machine learning techniques used were Support Vector Machines (SVMs), Linear Discriminant Analysis (LDA), Classification Trees, K-Nearest Neighbor (K-NN) and ensembles of trees and LDA. The results demonstrated a decline of up to 15% of the classification error with respect to the baseline value.

A method was proposed to recognize the fruits by Seng, W. C., and Mirisaee, S. H. (2009). The proposed methodology made use of three classes of features namely, shape-based, color-based and size-based features. K-Nearest Neighbor technology was used to recognize the fruits that demonstrated an accuracy of 90%.

An automatic identification system for fruits was developed by Aibinu, A. M. et al. (2011). The system was developed to identify three of the fruits viz. apple, mango and banana. The RGB features were used for the purpose of segmentation. Fourier descriptors and spatial domain features were used to recognize the shape of the fruits. Artificial Neural Networks (ANNs) were used for the purpose of training. Results of testing showed an accuracy of 99.1%.

6. MAJOR ISSUES AND RESEARCH GAPS ADDRESSED BY THE RESEARCHERS

The aforementioned study on the previous research works considered grading and quality assessment of various fruits such as apple, banana, cashew, dates, jatropha, nectarine, oil palm, orange, papaya, raisings, rapeseed, strawberry and tomato. The works mentioned here were successfully able to overcome many of the problems associated with post harvest handling issues such as labor requirements, fatigue of workers, subjective nature of the grading process leading to inconsistency, higher costs etc. Following are some of the glimpses of using machine vision in agricultural imaging tasks:

- Analysis of the images captured under laboratory conditions was done efficiently.
- Cost effective solutions are proposed for almost every fruit bearing commercial value.
- Time efficient solutions are given for few of the fruits.
- Efficient image processing algorithms are developed to sort / grade fruits and other horticultural produce.
- Modified machine learning algorithms are proposed and used efficiently to sort / grade few of the fruits.
- Optimization techniques are employed efficiently to sort / grade few of the fruits.
- Most of the solutions proposed are automated in nature.

7. ISSUES AND RESEARCH GAPS TO BE ADDRESSED

The use of machine vision to effectively address most of the post harvest handling issues of fruits found to be successful in majority of the cases. However, there are still thirst areas in which there is a need and scope to expand the same in order to address the issues as described below:

1. **Differences in production areas:** The research works carried out by the researchers considered a set of sample fruits' dataset produced from the same production area. However, there is a vital need for an unified model to grade and assess a particular fruit and/or fruit cultivar grown from different production areas.
2. **Seasonal Fluctuations in Grading Criteria:** Although there are standard guidelines available for grading different varieties of fruits, the seasonal fluctuations need to be considered. This is because, the growth and characteristics of the fruits change as per the season in which it is grown and harvested. Hence there is a need for an unified model to grade a fruit and/or fruit cultivar that can consider all seasonal fluctuations.
3. **Need for Sorter/Graders for Domestic Purpose:** Most of the research works concentrated on grading and quality assessment of fruits with commercial importance and export potential. However, there is a crucial need for developing cost effective and esthetic solutions to grade almost every fruit for the purpose of domestic transportation and distribution. Such solutions can enhance the monetary stability of growers/farmers.

8. CONCLUSION

Horticulture is the science and art of cultivating and marketing flowers, fruits, vegetables, ornamental plants and nuts. Fruits play an insignificant role in human nutrition. Production and marketing of fruits play an essential role in maintaining the monetary stability of the growers/farmers there by contributing to the financial growth of a country. Grading and quality assessment is the first and foremost step in post harvest handling of fruits as it decides the true market value of the commodity. Researchers round the clock are applying sincere efforts in inculcating Information and Communication Technologies in order to automate the process of grading and quality assessment of almost every fruit. In view of this, a comprehensive review of the research works carried out in the field of automated sorting/grading/quality assessment of fruits has been outlined in the present chapter. The major issues and research gaps addressed by the researchers are presented. At the end of the discussion, the issues and research gaps to be addressed are also outlined that serves as motivation and escorting material, for the researcher working in this arena, in developing the machine vision applications for various fruits and hence other horticultural produce.

REFERENCES

Aibinu, A. M., Salami, M. J. E., Shafie, A. A., Hazali, N., & Termidzi, N. (2011, January). Automatic fruits identification system using hybrid technique. In *Electronic Design, Test and Application (DELTA), 2011 Sixth IEEE International Symposium on* (pp. 217-221). IEEE. 10.1109/DELTA.2011.47

Arefi, A., Motlagh, A. M., Mollazade, K., & Teimourlou, R. F. (2011). Recognition and localization of ripen tomato based on machine vision. *Australian Journal of Crop Science*, *5*(10), 1144.

Clement, J., Novas, N., Gázquez, J. A., & Manzano-Agugliaro, F. (2012). High Speed Intelligent Classifier Of Tomatoes By Colour, Size And Weight. *Spanish Journal of Agricultural Research*, *10*(2), 314–325. doi:10.5424jar/2012102-368-11

Collins, R. J. (2009). *Value Chain Management And Postharvest Handling: Partners In Competitiveness*. Academic Press.

Dadwal, M., & Banga, V. (2012, October). Estimate Ripeness Level of fruits Using RGB Color Space and Fuzzy Logic Technique. *International Journal of Engineering and Advanced Technology*, *2*(1), 225–229.

Dang, H., Song, J., & Guo, Q. (2010, August). A Fruit Size Detecting And Grading System Based On Image Processing. In *Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2010 2nd International Conference on* (Vol. 2, pp. 83-86). IEEE. . 12010.1109/IHMSC.2010.120

Department of Horticulture. (2014). *Post Harvest Technology of Horticultural Crops. HISAR*. Haryana, India: CCS Haryana Agricultural University.

Effendi, Z., Ramli, R., Ghani, J. A., & Rahman, M. N. A. (2009, November). Pattern Recognition System Of JatrophaCurcas Fruits Using Back Propagation. In *Signal and Image Processing Applications (ICSIPA), 2009 IEEE International Conference on* (pp. 58-62). IEEE. doi: 10.1109/ICSIPA.2009.5478719

Font Calafell, D. (2014). An Image Processing Method For In-Line Nectarine Variety Verification Based On The Comparison Of Skin Feature Histogram Vectors. *Computers and Electronics in Agriculture*, *102*, 112–119. doi:10.1016/j.compag.2014.01.013

Horticulture Statistics Division, Department of Agriculture, Cooperation & Farmers Welfare, Ministry of Agriculture & Farmers Welfare. (2017). *Horticulture Statistics At A Glance 2017*. New Delhi: Government of India Controller of Publication Indian Council of Agricultural Research. Retrieved Jan 13, 2019, from https://icar.org.in/content/horticultural_ division.

Iraji, M. S., Tosinia, A., Azad, I., & Branch, A. (2011). Classification Tomatoes On Machine Vision With Fuzzy The Mamdani Inference, Adaptive Neuro Fuzzy Inference System Based (Anfis-Sugeno). *Australian Journal of Basic and Applied Sciences*, *5*(11), 846–853.

Jamil, N., Mohamed, A., & Abdullah, S. (2009). Automated Grading of Palm Oil Fresh Fruit Bunches (FFB) using Neuro-Fuzzy. 2009 *International Conference of Soft Computing and Pattern Recognition*, 245-249. Malacca.10.1109/SoCPaR.2009.57

Kader, A., & Rolle, R. (2004). *The Role Of Post-Harvest Management In Assuring The Quality And Safety Of Horticultural Produce*. Retrieved Jan 13, 2019, from http://www.fao.org /docrep/007 /y5431e/ y5431e00.htm#Contents

Kavdır, I., & Guyer, D. E. (2008). Evaluation Of Different Pattern Recognition Techniques For Apple Sorting. *Biosystems Engineering*, *99*(2), 211–219. doi:10.1016/j.biosystemseng.2007.09.019

Kondo, N., Ahmad, U., Monta, M., & Murase, H. (2000). Machine Vision Based Quality Evaluation OfIyokan Orange Fruit Using Neural Networks. *Computers and Electronics in Agriculture*, *29*(1-2), 135–147. doi:10.1016/S0168-1699(00)00141-1

Kurtulmuş, F., & Ünal, H. (2015). Discriminating Rapeseed Varieties Using Computer Vision And Machine Learning. *Expert Systems with Applications*, *42*(4), 1880–1891. doi:10.1016/j.eswa.2014.10.003

Lee, D. J., Chang, Y., Archibald, J. K., & Greco, C. G. (2008, August). Color Quantization And Image Analysis For Automated Fruit Quality Evaluation. In *Automation Science and Engineering, 2008.CASE 2008. IEEE International Conference on* (pp. 194-199). IEEE. 10.1109/COASE.2008.4626418

Leemans, V., & Destain, M. F. (2004). A Real-Time Grading Method Of Apples Based On Features Extracted From Defects. *Journal of Food Engineering, 61*(1), 83-89. doi: (03)00189-4 doi:10.1016/S0260-774

Leemans, V., Magein, H., & Destain, M. (2002, July). On-line Fruit Grading according to their External Quality using Machine Vision. *Biosystems Engineering*, *83*(4), 97–404.

Liming, X., & Yanchao, Z. (2010, September). Automated Strawberry Grading System Based On Image Processing. *Computers and Electronics in Agriculture*, *71*(1), S32–S39. doi:10.1016/j.compag.2009.09.013

Lorestani, A. N., Omid, M., Bagheri-Shooraki, S., Borghei, A. M., & Tabatabaeefar, A. (2006). Design And Evaluation Of A Fuzzy Logic Based Decision Support System For Grading Of Golden Delicious Apples. *International Journal of Agriculture and Biology*, *8*(4), 440–444.

May, Z., & Amaran, M. (2011). Automated ripeness assessment of oil palm fruit using RGB and fuzzy logic technique. In M. Demiralp, Z. Bojkovic, & A. Repanovici (Eds.), *Proceedings of the 13th WSEAS international conference on Mathematical and computational methods in science and engineering (MACMESE'11)* (pp. 52-59). Catania, Sicily, Italy: World Scientific and Engineering Academy and Society (WSEAS).

Mustafa, N. A., Fuad, N. A., Ahmed, S. K., Abidin, A. Z., Ali, Z., & Yit, W. B. (2008). Image Processing Of An Agriculture Produce: Determination Of Size And Ripeness Of A Banana. *2008 International Symposium on Information Technology*, 1-7. doi: .463163610.1109/ITSIM.2008

Mustafa, N. B. A., Ahmed, S. K., Ali, Z., Yit, W. B., Abidin, A. A. Z., & Sharrif, Z. A. M. (2009, November). Agricultural produce sorting and grading using support vector machines and fuzzy logic. In *2009 IEEE International Conference on Signal and Image Processing Applications* (pp. 391-396). IEEE. doi: .2009.547868410.1109/ICSIPA

Nakano, K. (1997). Application Of Neural Networks To The Color Grading Of Apples. *Computers and Electronics in Agriculture*, *18*(2-3), 105–116. doi:10.1016/S0168-1699(97)00023-9

Narendra, V. G., & Hareesha, K. S. (2011). Cashew Kernels Classification using Colour Features. *International Journal of Machine Intelligence*, *3*(2), 52–57.

Omid, M., Abbasgolipour, M., Keyhani, A., & Mohtasebi, S. S. (2010). Implementation of an efficient image processing algorithm for grading raisins. *International Journal of Signal and Image Processing*, *1*(1), 31–34.

Riyadi, S., Ishak, A. J., Mustafa, M. M., & Hussain, A. (2008, May). Wavelet-Based Feature Extraction Technique For Fruit Shape Classification. In *Mechatronics and Its Applications, 2008.ISMA 2008. 5th International Symposium on* (pp. 1-5). IEEE. 10.1109/ISMA.2008.4648858

Rocha, A., Hauagge, D. C., Wainer, J., & Goldenstein, S. (2010). Automatic Fruit And Vegetable Classification From Images. *Computers and Electronics in Agriculture, 70*(1), 96-104. doi: .09.002 doi:10.1016/j.compag.2009

Rokunuzzaman, M., & Jayasuriya, H. P. W. (2013). Development of a low cost machine vision system for sorting of tomatoes. *Agricultural Engineering International: CIGR Journal*, *15*(1).

Seng, W. C., & Mirisaee, S. H. (2009, August). A new method for fruits recognition system. In *Electrical Engineering and Informatics, 2009.ICEEI'09.International Conference on* (Vol. 1, pp. 130-134). IEEE. doi: 10.1109/ICEEI.2009.5254804

Unay, D., & Gosselin, B. (2005a). Thresholding-Based Segmentation And Apple Grading By Machine Vision. *Signal Processing Conference, 2005 13th European*, 1-4.

Unay, D., & Gosselin, B. (2005b). Artificial Neural Network-Based Segmentation And Apple Grading By Machine Vision. In *Image Processing, 2005.ICIP 2005.IEEE International Conference on* (Vol. 2). IEEE.10.1109/ICIP.2005.1530134

Xiaobo, Z., Jiewen, Z., & Yanxiao, L. (2007, November). Apple Color Grading Based On Organization Feature Parameters. *Pattern Recognition Letters*, *28*(15), 2046–2053. doi:10.1016/j.patrec.2007.06.001

Yamamoto, K., Ninomiya, S., Kimura, Y., Hashimoto, A., Yoshioka, Y., & Kameoka, T. (2015, January). Strawberry Cultivar Identification And Quality Evaluation On The Basis Of Multiple Fruit Appearance Features. *Computers and Electronics in Agriculture*, *110*, 233–240. doi:10.1016/j.compag.2014.11.018

Yang, L., Dickinson, J., Wu, Q. J., & Lang, S. (2007, December). A Fruit Recognition Method For Automatic Harvesting. In *2007 14th International Conference on Mechatronics and Machine Vision in Practice* (pp. 152-157). IEEE. 10.1109/MMVIP.2007.4430734

Chapter 10
Optimized Data Mining Techniques for Outlier Detection, Removal, and Management Zone Delineation for Yield Prediction

Roopa G. M.
Bapuji Institute of Engineering and Technology, India

Naveen Kumar K. R.
Bapuji Institute of Engineering and Technology, India

Arun Kumar G. H.
Bapuji Institute of Engineering and Technology, India

Nirmala C. R.
Bapuji Institute of Engineering and Technology, India

ABSTRACT

Enormous agricultural data collected using sensors for crop management decisions on spatial data with soil parameters like N, P, K, pH, and EC enhances crop growth for soil type. Spatial data play vital role in DSS, but inconsistent values leads to improper inferences. From EDA, few observations involve outliers that deviates crop management assessments. In spatial data context, outliers are the observations whose non-spatial attributes are distinct from other observations. Thus, treating an entire field as uniform area is trivial which influence the farmers to use expensive fertilizers. Iterative-R algorithm is applied for outlier detection to reduce the masking/swamping effects. Outlier-free data defines interpretable field patterns to satisfy statistical assumptions. For heterogeneous farms, the aim is to identify sub-fields and percentage of fertilizers. MZD achieved by interpolation technique predicts the unobserved values by comparing with its known neighbor-points. MZD suggests the farmers with better knowledge of soil fertility, field variability, and fertilizer applying rates.

DOI: 10.4018/978-1-5225-9632-5.ch010

INTRODUCTION

Agriculture plays a major role in overall socio-economic development of a Country but nowadays agricultural contribution towards the economic growth is steadily decreasing with country's wide ranged economic growth. But still, agriculture is demographically the largest economic sector. In most of the countries, agriculture is a combination of traditional and modern farming techniques. Current agricultural practices are neither economically nor environmentally good. The average size of land holdings is very small and is subjected to fragmentation where such small holdings are repeatedly over-manned, which results in low productivity. Thus, Precision Agriculture as a novel approach is suggested to tackle the above mentioned problems and it also creates new opportunities for data intensive science in the multi-disciplinary agro-environmental domain (Rob Lokers et al., 2016).Precision Agriculture is a generic term which includes the knowledge of plant and animal science and practical application of procedures (machines, treatments, tools, supplies). In broad sense Precision Farming or Precision Agriculture can be efficiently defined as the use of information technology to improve the decision making process in agricultural production. The data chain interacts with farm processes and farm management processes through various decision making processes in which information plays an important role (Sjaak Wolfert et al.,2017).Precision Agriculture practice contributes in improving the efficiency of production and decreasing environmental impact. Since the beginning of Precision Agriculture technology epoch, patterns of crop variability have been considered crucial for variable rate nutrient management. It also acts as a unique crop production business which is dependent on many climate and economy factors soil, climate, cultivation, irrigation, fertilizers, temperature, rainfall, harvesting, pesticide weeds and other factors (Jharna M. et al., 2017).

Components Involved in Precision Agriculture

Precision agriculture is information intensive field as it requires essential layers of data in order to provide the necessary information for precise decision making. The initial stage of implementing Precision Agriculture involves the process of collecting geo-reference crop yield data which is referred as yield-mapping that results in a document which represents the spatial pattern of crop yield and variables which are present during the plantation period. Yield maps as evidence examines sufficient farm spatial variability to implement site specific nutrient/fertilizers management. For such sufficient spatial variability soil sampling process is required to characterize soil properties to formulate management zones for application inputs. Further, to process and interpret the soil information the quantitative methods are used in digital soil mapping phase. For such given application inputs the general decision making treats enter field area as homogeneous that are unique to each zone in management zone phase . Next, Variable-Rate-Technology (VRT) permits the agriculture input like fertilizers/ pesticides and herbicides applied on-the-go all over the field at suitable rates according to the application map. Finally, in site-specific crop management phase the resource applications are matched with soil attributes and crop requirement for better yield productivity.

Outliers in Spatial Data Context

With the growth and usage of spatial data in precision agriculture, the challenges arise with the need to retrieve the useful spatial information which mainly emphasizes for applying better data pre-processing techniques. While obtaining the spatial yield datasets from remote sensor/ yield monitoring sensors and GPS various random and systematic errors may occur due to natural topographic conditions and measurement errors. Such errors have huge impact on yield measurements that creates unrealistic measurements and inaccurate inferences. In order to gain better understanding of spatial data information such errors should be removed from the crop yield dataset (Miguel et al., 2014 and Qiao Cai et al., 2013). Detection or removal of errors in spatial yield data has significant importance in site-specific crop management. In spatial data context, the observations whose non spatial attribute deviates remarkably from other observations within its spatial proximity are termed as outliers. Outlier analysis is carried out to identify abnormal activities (B. A. Sabarish et a., 2018) that comprises a small portion of the whole dataset and reside in small clusters in sparse region and behave differently relative to the majority of the normal data (Yuan Wang et al., 2016; Nita M. et al., 2015).

Role of Management Zone Delineation for Yield Productivity

Precision agriculture can also be termed as site specific management with its basic principle support for implementing in a spatial scale smaller than that of an entire field. Number of information sources is used to delineate subfield management zones for site specific crop management where there is lot of data available which contains information about the land holdings. The available data with its soil properties help the farmers to gain yield productivity. Based on the biological valid assumption, a large area of field contains wide spatial imbalances in soil types and nutrients availability. When such variations are considered, the results end in a loss of productivity. An important task to increase the productivity is by adopting a concept called as Management Zone Delineation which divides the agricultural field into homogeneous sub-fields or zones based upon the soil parameters (F. Guastaferro et al., 2010).

Site specific methods results in imbalanced field management which impacts on the crop productivity whereas the critical task in convention agriculture is base fertilization. Due, to this farmer treats the entire field as uniform area which forces them to use expensive resources like fertilizers/pesticides. Understanding spatial variation within a field is essential for site-specific crop management, which requires the delineation of management areas (Córdoba. M et al., 2013).In heterogeneous environment it is an acute task to locate which field portion must be considered and the quantity of fertilizers and pesticides required to gain high yield productivity a concept of Management Zone Delineation has to be adopted.

In some regions, the agricultural fields for small plot where the concept of Management Zone Delineation can be applied efficiently. The use of homogeneous management zones has demonstrated good potential for the site specific crop management. Figure 1 shows the division of farms into small plots.

Farms in some regions give better yields whereas some not and yields for some agricultural commodities are also low. Authors have conclude that crop yields very significantly so Management Zone Delineation gives the farmer a better understanding of soil fertility variability within a field and how to adjust fertilizer application rates.

Objective

The main objective of the work is to provide better suggestions and understanding of the farm-area which relies on data-driven decision support system for yield prediction and Site Specific precise farm management. The work extends to determine yield-potential for every yield and fertile zone which helps to generate site-specific recommendations that enhances to improve application of fertilizers for efficient crop-management. Thus, Profitability in smart-farming can be achieved by suggesting the farmers for optimal usage of agro-inputs and fertilizers /nutrients, seeds, natural energy sources, pesticides, herbicides, chemicals, etc.,

Figure 1. Small-plot areas

BACKGROUND THEORY

Precision Agriculture

Precision agriculture also known as “Smart-farming” involves the usage of many principal-technologies like Communication-systems, positioning technologies-telematics, hardware / software applications, sensing technologies, data mining techniques, application of big-data discussed by Jharna Majumdar et al.(2017) and data-analytics solutions proposed by Rob Lokers et al.(2016) to optimize the crop-yield performance. The expertise and regulation necessitates GPS technology, image-data manipulation, computer-based data analysis, specific mathematical-models and Scientific-frameworks to manipulate, analyze and interpret the data. Even-though, broader endorsement of the technical infrastructure J.W. Kruize et al. (2016) encouraged the deployment of modern tools and technologies, a typical farmer eventually faces hurdles in the adoption of successive follow-ups due to the possibility of inadequacy in local technical guidance and lack of knowledge & expertise, limits on infrastructure with a risk of insufficient return on the investment explored by Sjaak Wolfert et al. (2017). The profitability in smart-farming can be incurred by the reduced usage of agricultural-inputs such as fertilizer / nutrients, seeds, energy, pesticides-herbicides, chemicals, water, etc, but the goal of most farming-practices in agriculture is to approach a management and assistance to optimize agricultural-production thereby achieving the profitability.

Spatial-Outliers

Nowadays, Smart-farming is highly technological-oriented. The Smart-agriculture field constructively utilizes the strength that can be transformed in to the actions by big-data. Software, Analytical-skills are used to develop algorithms to pull vital inferences out of critical data. Information-driven management techniques are playing a major role, where, Spatial-Data is the building block of Precision farming. Yield prediction is a usual issue but a substantial task where a farmer expects a specific amount of yield based on both his past knowledge about particular farm-fields and data collected to manage the fields during regular farming operations using various sensors. The Data collected from sensing technologies usually are fine-scale, carries auto-correlated or highly-correlated spatial information about massive area. The decision support system presented by Pornchaiet al. (2014) uses data mining techniques to manipulate data which is vital in yield prediction should determine both temporal and spatial variability associated with a farm field in order to manage site-specific precise farming approach. The variability in temporal / spatial data may be contributed by several factors during data acquisition. Understanding the result of these individual consideration can only be measured and managed using statistical analysis of the data. Hence, Detection of errors / outliers in accumulated data is important for site-specific crop management and numerous statistical-techniques expressed in this regard. In a spatial-data context, outlier detection techniques are distinguished into two sets of Spatial-outlier tests: Graphical and Quantitative discussed by Peter Chu Su (2011). Quantitative methods work on statistical tests which distinguish spatial outliers from the rest of the dataset. These include the linear regression (Scatterplot) and Spatial Statistic Z George Rub (2010). Swamping (failing to identify outliers) and Masking (mistaking clean observations for outliers) Robert Serflinget al. (2012) considered these effects to eliminate/minimize by spatial outlier detection techniques and S-outlier-algorithms, such as Iterative-Z, Iterative-R and Median-Z can be served for these purpose.

B. A. Sabarish et al. (2018) and Aminaet al. (2014) suggested the trajectories can be detected using convex hull algorithms. The trajectories are then classified as normal or outlier using the ray casting algorithm. The main contribution of the work was to identify outlier trajectories using boundary method and their classification was done based on the constructed boundary. LTI Information Treatment Laboratory, Morocco, the authors have proposed a hybrid approach between distance-based and density-based approaches which has presented a new approach for detecting outliers by introducing the notion of object's proximity.

Management-Zone-Delineation

Asim Biswas et al. (2013) suggested various models to optimize a specific soil-parameter differently and which indicated the uncertainly involved with the soil-parameter. The uncertainty of the parameter value makes the spatial-prediction processes very difficult/uncertain. Their work explained the average procedures used to generate the parameter-value and reduce the model soil-parameter uncertainty. The work presented semi-variogram as the fitted model with four commonly used mathematical-models (Gaussian, spherical, linear plateau and exponential).

Córdoba, M et al. (2013), studied approaches for management-zones which divides the field-area into sub-fields with minimal soil-parameters variability that has maximum heterogeneity of soil-conditions and topography, so that the management-zones lead to the similar yield potentially. The study carried by Zhang X et al. (2016), evaluated the farmers experience in delineating the zones. The work used the evaluated data, sand, clay, soil-penetration and the spatial correlation matrix was generated that influenced the yield-productivity. Further, the work involved in selecting the interpolated data and delineating the zones by Fuzzy C-means clustering approach with Software-Delineating-Management-Zones. The work presented by George et al. (2010), focused on the procedures that are suitable huge multi-variant datasets. They adopted Fuzzy k-means for delineating the zones in a multi-variant context to carry-out the analysis. The issues identified lead to the extension of PCA-algorithms and KM-sPC resulted in high yield differences among the delineating classes and exhibited small within class yield-variance.

The intuitive method proposed by Nahuel et al. (2015) supported site-specific crop-management of N by adjusting the fertilizer-rates based on the soil-characteristics. The work aimed at ensuring whether delineating the zones within the field-area enhances the NUE in wheat. The suggested method afforded the opportunity of applying variable-rate of n-fertilizers and also leaded to minimize the risk of pollution due to excessive application of fertilizers.

Summary of Literature Review

In the context of Precision Agriculture and data-mining, numerous spatial outlier detection techniques have been proposed which supersedes existing techniques. But it is unclear that whether the new algorithms are better than the existing methods. In addition to this, the comparisons of spatial outlier detection techniques are remained unexplored. Spatial–Outlier algorithms are applied for implementation and analyzing how they are efficient in terms of minimizing the swamping and masking effects. Some of the machines learning algorithms are explored for outlier detection and removal process. It is observed that, most popular statistical spatial outlier algorithms used on the spatial data to detect the spatial outliers introduced are based on the principle that compares the attribute value at each location against an aggregate function that summarizes the neighborhood attribute values. Further it is observed that currently, many spatial outlier detection techniques are available but there is no knowledge about which spatial outlier detection technique is better. The Problem of finding irregular feature in spatial data to deal with masking and swamping effects need to be explored and considered MZD issue.

Here, the delineated management-zones rely on the usage of particular applications. Where different approaches are inherently-subjective and no-generic approach would fit the applications. Thus, there is no absolute efficient measure to check the results of management-zones delineated. In such case, it is necessary to enhance the understanding of the study-area and get insights over the available datasets. Hence, MZD should support for the exploration of the spatial-data rather than the fixed clustering. Applying various techniques like PCA, linear data combinations and further data-transformations leads to crisps-zones and straight forward-computations which results in loss of understanding of the newly created variables. The other issue explored is the usage of spatial-methods. The application of non-spatial methods (k-means/fuzzy c-means) generates continuous-zones that are sometimes smooth/fixed and finally gives unstable-results. The above problem can be solved by changing the k/c parameter leading to various management-zones with high spatial-layouts.

METHODOLOGY PRESENTED

The methodology presented in this research work is discussed in Figure 2. The acquired data with the soil-chemical/physical properties is gathered by simulation/prediction and further used for crop-management assessment.

Phase One: Experimental Site

The study is carried out at various farm fields of Davangere region during 2016 to 2018 from Taralabalu Krishi Sanshodhana Kendra. The study area consists of co-ordinate points in every experimental site which are identified with a GPS receiver. The boundary overview of the study area is shown in Figure 3. The soil samples are manually collected where standard procedures are applied to determine the soil properties like organic matter (N,P,K). With this generated data three distinct datasets are made available namely crop dataset, map dataset and plot dataset in association with Krishi Sansadhan Kendra. The prepared dataset contains 685 observations for different villages within Davangere District Jurisdiction.

The standard percentage of nitrogen, Phosphorus, potassium and pH content required by crops which varies from nitrogen ranges from 48 to 236 kgha-1, Phosphorus ranges from 8 to 78 kgha-1, Potassium ranges from 9231 to KG kgha-1 and pH ranges from 6.7 to 7.7 kgha-1.

Phase Two: Exploratory Data Analysis

Exploratory Data Analysis (EDA) is considered to be a best practice to explore the data with the help of statistical / visualization techniques to get a sight into significant exposure of the data for future analysis. Exploratory Data Analysis is regarded as a critical-process that initially investigates the data in order to find patterns, to locate mistakes, missing values, anomalies and mapping-out the underlying structure and nature of the data with the help of graphical representations and valid summary-statistics. Uni-variate and multivariate visualizations are performed using statistical programming packages from R, where uni-

Figure 2. Methodology presented

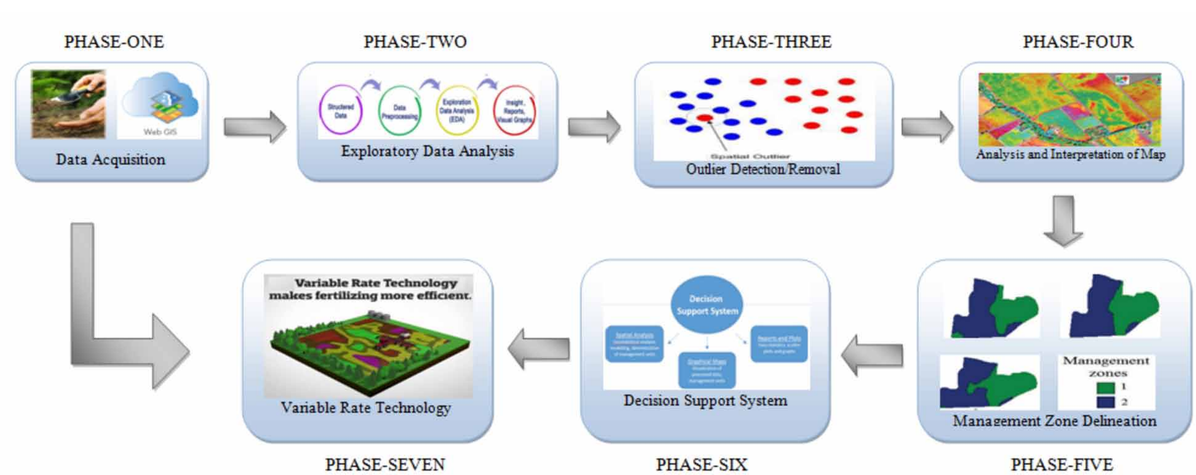


Figure 3. Boundary of study area

variate visualization gives summary-statistics for individual variable in the raw data-set and multivariate visualization provides information to know the interactions between various fields from the dataset.

Phase Three: Spatial-Outlier Detection/Removal

Spatial data is the major building block of Precision farming whether the data comes from soil sampling or any satellite connection signal that indicates both temporal and spatial variability. Within the farm field search variability is contributed by several factors understanding the effect of each factor can only be measured and managed using statistical analysis of the data.

In statistics, few observations deviate from other observations are regarded as outliers. An outlier may arise due to variability in measurement which indicates the experimental errors. This may lead to catastrophic issues in statistical analysis. Thus, outliers indicate inconsistent data erroneous procedures and Zone area where certain theories might be invalid. For any acquired spatial data with large soil samples a few number of outliers are expected (Robert S. et al. and Shashi S, et al., 2012). Outliers are the exceptional observations which comprises of samples with max-min values. But some samples with max and min values do not exhibit the nature of outlyingness as they may not usually distinct from normal observations. Concluding any observation as an outlier is ultimately a subjective exercise (AminaDik et al., 2014).

Various statistical methods for outlier detection/removal in spatial context are suggested which involves quantitative and graphical approaches (George-Rubet al., 2010 and Peter Chu Su,2011). Graphical spatial outlier methods work on the visualization of spatial data to detect outliers. Approaches like Moran-Scatterplot and Variogram-Cloud are graphical based outlier detection approaches and quantitative techniques such as linear-regression, spatial-statistics are mainly based on statistical tests that distinguish outliers from the other consistent values of the dataset. Further masking and stamping are the effects considered to eliminate.

Swamping is failing to identify outliers and masking is mistaking clean observation for outliers. There are three spatial outlier algorithms to reduce masking and swamping effects. Iterative-R, Iterative-Z and Median-Z algorithm.

Phase Four: Management-Zone-Delineation

Smart-Farming involves collecting and monitoring agronomic information/practices to provide precise resources and needs to sub-fields rather than the average resources for the entire field. Early days of farming involves varying field crop inputs which are based on farmer assumptions, survey on soil-maps, soil sample testing, direct soil observation and weather conditions while analyzing the field variability. Many agronomic studies with farm field experiments have revealed the significance of site-specific characteristics that has a great impact on crop yield productivity. Such site specific resource applications of crop-input can be accomplished by dividing the entire field into small plots referred as "Homogeneous Management Zones" given in Figure 1.

Agronomics have identified the limitations for grid soil-sampling by considering additional site specific characteristics for variable-rate to crop input strategies like historical/intuitive, qualitative and quantitative factors which are stable/dynamic over the years. In this regard, farm-management zones can be conceptualized as "A sub-area of a farm-field which exhibits homogeneous combination of crop yield limiting aspects with an appropriate field allocation of site specific crop yield inputs. Thus, delineating of management-zones is a task of classifying the spatial-data variability for a given specific field. Hence, the strategies for delineating management-zones must follow the true-cause and the effective relationship between the site specific characteristics and the crop yield.

The spatial information which is a significant factor for defining the management-zones should be densely/continuously sampled, quantitative, stable over time and directly relative to crop yield. The primary factor for accurate/profitable applications of crop-inputs the uniform variable rate resource directly has a great impact on crop growth and yield productivity.

Management-Zone-Delineation can be accomplished by interpolation method called kriging. kriging implicates an investigation of spatial data behavior and supports the prime interpolation technique to generate the output soil surface. It employs statistical model based on auto-correlation which refers to the statistical relationship within the measured soil-points. The process involves exploratory statistical-data analysis, variogram modeling, soil surface creation and to explore the variance of the soil surface. Moreover, management zones with rectangular shapes are more applicable for farming in underdeveloped areas because farmers can easily apply these management zones to reduce fertilizer input, labor costs, and environment waste without using advanced agriculture machinery (Xiaohu Zhang et al., 2016).

Phase Five: Decision-Support-System (DSS) and Variable-Rate-Technology (VRT)

Decision Support System (DSS) plays a central role in precision farming. A set of statistical techniques are used for both studying the nature of data and analysis of spatial data. A farming manager can adopt best practices by a sound decision making with this field information. Finally, group of homogenous soil-sample clusters are formed and the soil physico-chemical variable rates are identified for these clusters to apply the precise crop-inputs based on the time/location to attain site-specific input rates using VRT. Thus, the effectiveness of these agricultural Decision-Support-System is reliant on the type of environmental, crop and other data that is collected by sensors which can be integrated into the DSS or interrogated further using data mining or other analysis techniques (Pornchai T. et al., 2014).

EXPERIMENTAL RESULTS, ANALYSIS AND INTERFACES

Experimental Setup

Schematic Representation

The Figure 4 depicts the Experimental setup for precision farm management system and the experimental setup for the boundary area is laid-out by deploying various farm-field sensors for spatial-data acquisition. Wireless-weather station with PC-Interface is responsible for exporting the data to the nearest data-server using Rasberry Pi-3 interface. The data mining algorithms identified in this work are implemented using R-Studio. The set of R-packages/interfaces discussed in Table 1 are applied. The user front-end is designed using Shiny-App.

Packages/Interfaces Used

R-Studio, R Programming-Language and important R-packages are used for implementing the presented work. The details of the packages/interfaces used are discussed in the Table 1.

Figure 4. Experimental set-up for precision farm management system

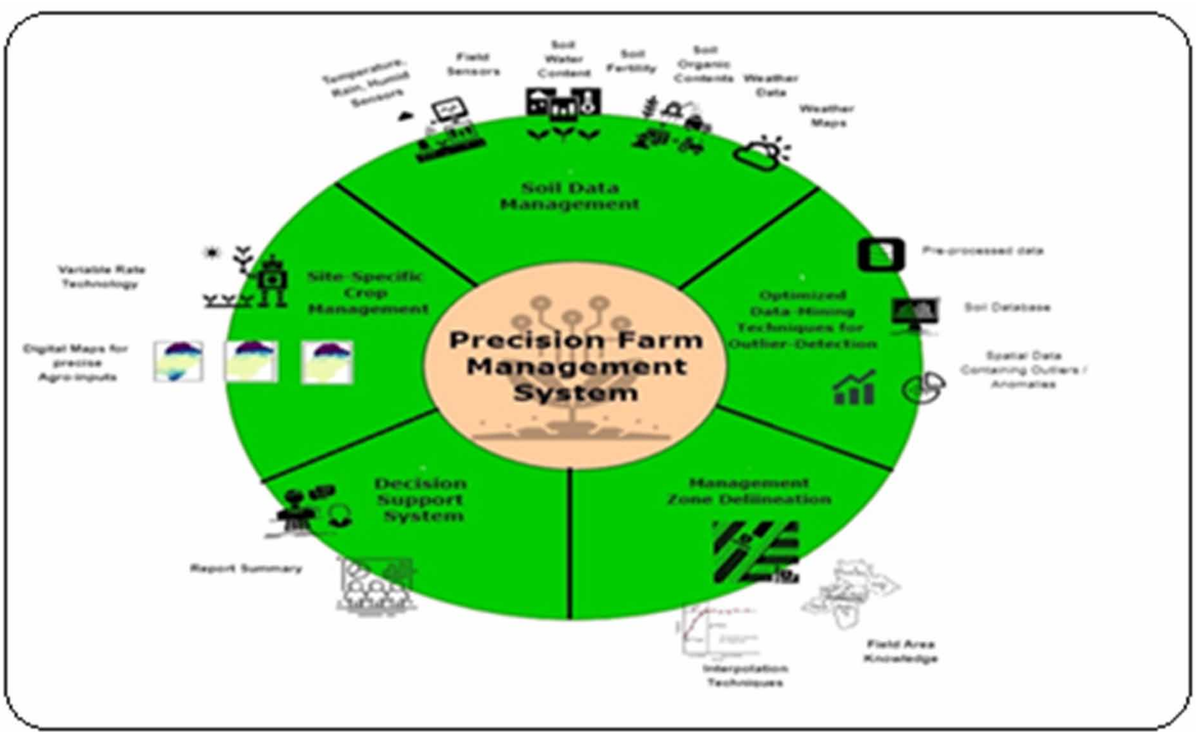


Table 1. Description of interfaces/packages used

Packages/ Interfaces	Description
S_p	Classes/methods for spatial-data, class-document which holds the spatial-location information for 2D/3D data.
Geo-R	Geo-Statistical analysis including likelihood-based, traditional and Bayesian-methods.
Gg-Map	Set of functions for visualizing spatial-data using Google-Maps.
Auto-Map	Performs automatic-interpolation to estimate the variogram by triggering gstat() method.
shiny-app	Interactive-web-applications that builds Automatic-reactive-binding Between inputs and outputs and extensive-prebuilt-widgets.
3D-Scatterplot	Plots a three dimensional (3D) point in three-dimensional-space.
raster	Reads/writes, manipulates, analyzes and models gridded spatial data
Ggplot2	A system to create graphics based on the Graphics-Grammar whichmaps variables to aesthetics.
rgl	Gives medium to high-level-functions for 3D-interactive-graphics.
sql-df	For running SQL-statements on data-frames.
leaf-let	Open-source Javascript library for interactive-maps
rs-connect	Deployment interface for shiny-applications.

Dataset Description

The experimental study is carried out at various cultivated farm location within Davangere region in the year 2017-18 for various crop seasons. The soil sample data with the following properties like organic matter like pH, nitrogen, potassium and phosphorus availability is collected in association with Taralabalu Krishi Samshodhana Kendra. The coordinate points for each of experimental area are located with GPS receiver with a satellite enabled signal. An overview of study boundary (Davangere-region) is given in Figure 5. With the available data, two datasets are created such as plotted dataset and crop dataset.

Crop Dataset

Table 2, shows the details of various crops with its minimum soil sample values (fertilizers) needed for high crop yield productivity. The map dataset is generated with 208 different geo-graphic locations which covers the entire Davangere District region. The latitude and longitude values for each region are retrieved using Google Map API.

Plotted Dataset

Figure 6, shows the description of plotted dataset for Davangere region. It consists of various soil sample values like n, p, k region coordinates where X and Y refers to longitude and latitude. The dataset consists of 685 observations for six taluks of Davangere District. The entire Davangere region lies within longitude of 14.44 and 14.48N and latitude of 7488 270 5.95E.Authors have collected the following soil sample values within Davangere region based on n, p, k and pH content.

Figure 5. Boundary area for Davangere-district

For the collected spatial dataset, it is examined that it consists of missing values, redundant values and irrelevant information. In order to analyze the soil characteristics and discover the patterns to spot anomalies Exploratory-Data-Analysis technique is suggested.

Exploratory-Data-Analysis (EDA)

Exploratory Data Analysis involves visualization techniques to interpret the nature of acquired spatial data. Various built-in R-packages like Plot-3D, Scatter-plot, ggplot2 and rgl are used to generate graphs which are visualized to explore the distribution/concentration of n, p, k, pH, EC values associated with the soil sample for the given dataset. The graph visualization shows various scattered observations over the plane and plots shows huge variations in the observations. The individual histograms or scatter-plots for acquired n,p,k values depicts the field regions where huge data resides and also focuses on anomalies.

Figure 7 and Figure 8 shows the distribution of n, p, k values plotted with nitrogen, phosphorus and potassium values on X-axis and their number of occurrences along Y-axis. It is observed that minimum and maximum values for ‘N’ is 8-230 kgha$^{-1,}$ ‘P’ is 4-78 kgha^{-1} ‘K’ is 4-324 kgha^{-1}. Using 50-130 kgha^{-1}, majority of phosphorus values are concentrated between 5-24 kgha^{-1} and majority of potassium values are concentrated between 4-110 kgha^{-1}.

Various scatter plots are plotted to visualize the relationship between the parameters. Boxplot for n,p,k values is generated to show the data distribution with its upper and lower quartiles and IQR(Inter-Quartile-Range). Figure 9 shows that for the acquired nitrogen values the median is closer to the third quartile where as for the phosphorus or potassium the sample are comparatively closely dense between the third quartile and the median and it is more scattered between the first quartile and the median. The upper-whisker of phosphorus is marginally longer than the lower one which specifies that the data is

Table 2. Crop dataset

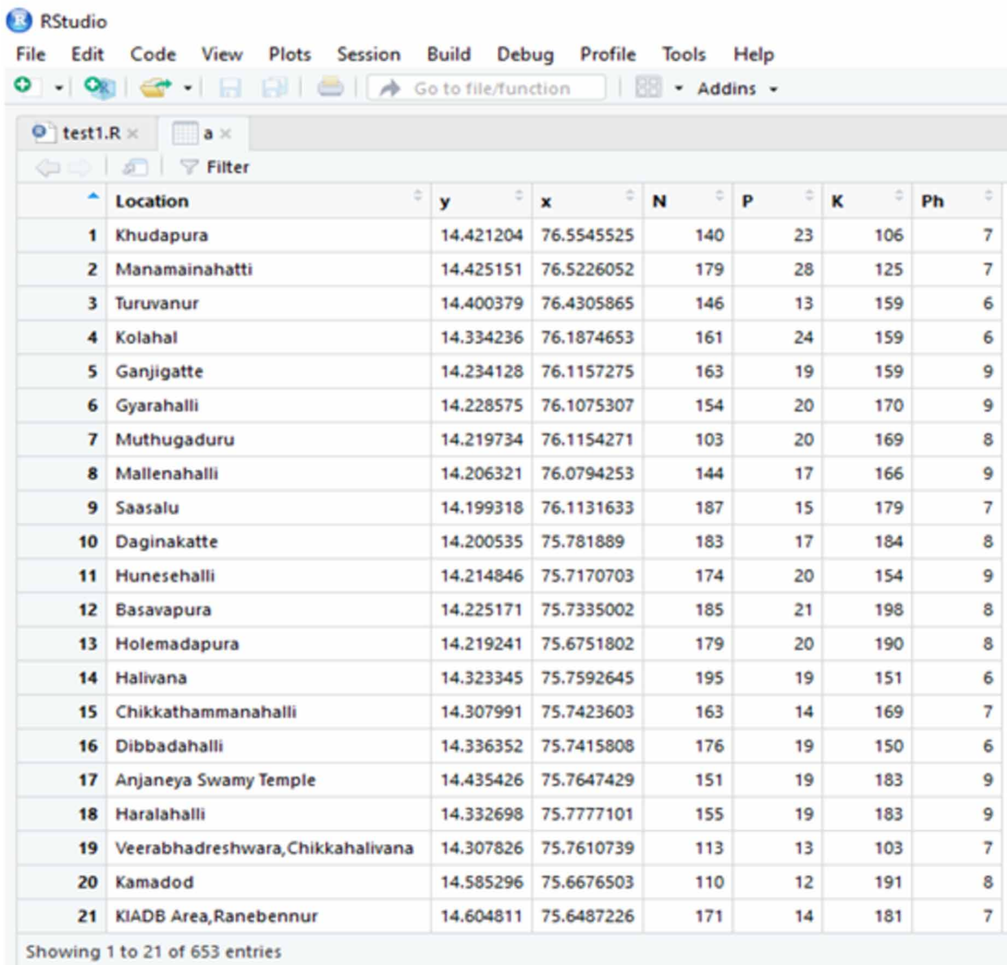
Crop	N	P	K	pH
Rice	80	40	40	5.5
Wheat	100	40	0	5.5
Jowar (Sorghum)	80	40	40	5.5
Barley (JAV)	70	40	45	5.5
Bajra (Pearl Millet)	90	30	0	5.5
Maize	80	40	20	5.5
Ragi (naachnnii)	50	40	20	5.5
Chickpeas (Channa)	40	60	80	5.5
French Beans (Farasbi)	90	125	60	5
Fava beans (Papdi - Val)	90	125	60	5
Lima beans (Pavta)	40	60	20	5
Cluster Beans (Gavar)	25	50	25	5
Soyabean	20	60	20	5.5
Peanuts	15	25	0	5.5
Black eyed beans (chawli)	20	60	20	5.5
Kidney beans	20	60	20	5.5

skewed towards right, whereas for potassium the lower whisker is slightly longer than the upper whisker which indicates that the data is scaled towards the left. The upper-lower whiskers for nitrogen lies at 60-240 kgha-1. Beyond this lower whisker there exists an outlier, similarly the upper lower whiskers for Phosphorus lies at 6-85 kgha-1. Here it is observed that few outliers beyond the upper-whisker exists. The upper-lower whiskers for potassium lies at 8-323 kgha-1 where it is observed that outliers exist beyond both whiskers. The box for potassium and nitrogen are slightly taller when compared to phosphorus whereas the upper-lower whiskers for nitrogen and potassium are longer than Phosphorus. From this it is explored that the values spread is slightly more for nitrogen and potassium and there exists several outliers among all the three soil parameters for the acquired spatial data. By applying boxplot on the spatial dataset the outliers can be effectively identified (Chang-Tien et al., 2003).

Implementation Details of Spatial-Data Outlier Detection/Removal

The authors have applied Iterative-R (ratio) spatial-data Outlier detection/removal algorithm on the pre-process dataset. The algorithm detects one spatial-outlier in each iteration. Further, the ratio point attributes (r-value) and their neighbor average-attribute values are computed for each point. Then, the ratio-point with the largest value (r-value) is identified as an outlier. Finally, the previously detected spatial-outlier attribute value is substituted with its neighbor average attribute-value.

Figure 6. Plotted dataset with minimum soil sample values



	Location	y	x	N	P	K	Ph
1	Khudapura	14.421204	76.5545525	140	23	106	7
2	Manamainahatti	14.425151	76.5226052	179	28	125	7
3	Turuvanur	14.400379	76.4305865	146	13	159	6
4	Kolahal	14.334236	76.1874653	161	24	159	6
5	Ganjigatte	14.234128	76.1157275	163	19	159	9
6	Gyarahalli	14.228575	76.1075307	154	20	170	9
7	Muthugaduru	14.219734	76.1154271	103	20	169	8
8	Mallenahalli	14.206321	76.0794253	144	17	166	9
9	Saasalu	14.199318	76.1131633	187	15	179	7
10	Daginakatte	14.200535	75.781889	183	17	184	8
11	Hunesehalli	14.214846	75.7170703	174	20	154	9
12	Basavapura	14.225171	75.7335002	185	21	198	8
13	Holemadapura	14.219241	75.6751802	179	20	190	8
14	Halivana	14.323345	75.7592645	195	19	151	6
15	Chikkathammanahalli	14.307991	75.7423603	163	14	169	7
16	Dibbadahalli	14.336352	75.7415808	176	19	150	6
17	Anjaneya Swamy Temple	14.435426	75.7647429	151	19	183	9
18	Haralahalli	14.332698	75.7777101	155	19	183	9
19	Veerabhadreshwara,Chikkahalivana	14.307826	75.7610739	113	13	103	7
20	Kamadod	14.585296	75.6676503	110	12	191	8
21	KIADB Area,Ranebennur	14.604811	75.6487226	171	14	181	7

Showing 1 to 21 of 653 entries

Figure 7. Histogram of ‘N’, ‘P’, ‘K’ Values

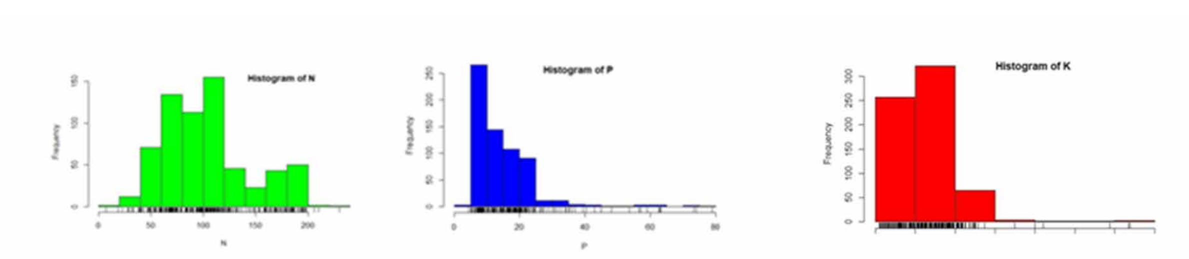


Figure 8. Scatter-plot of 'N', 'P' and 'K' values

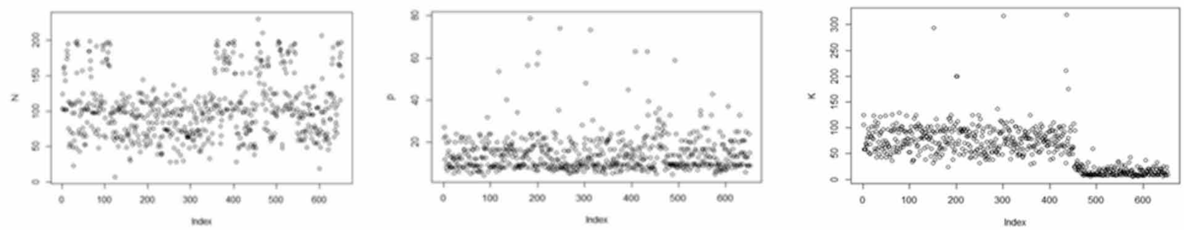
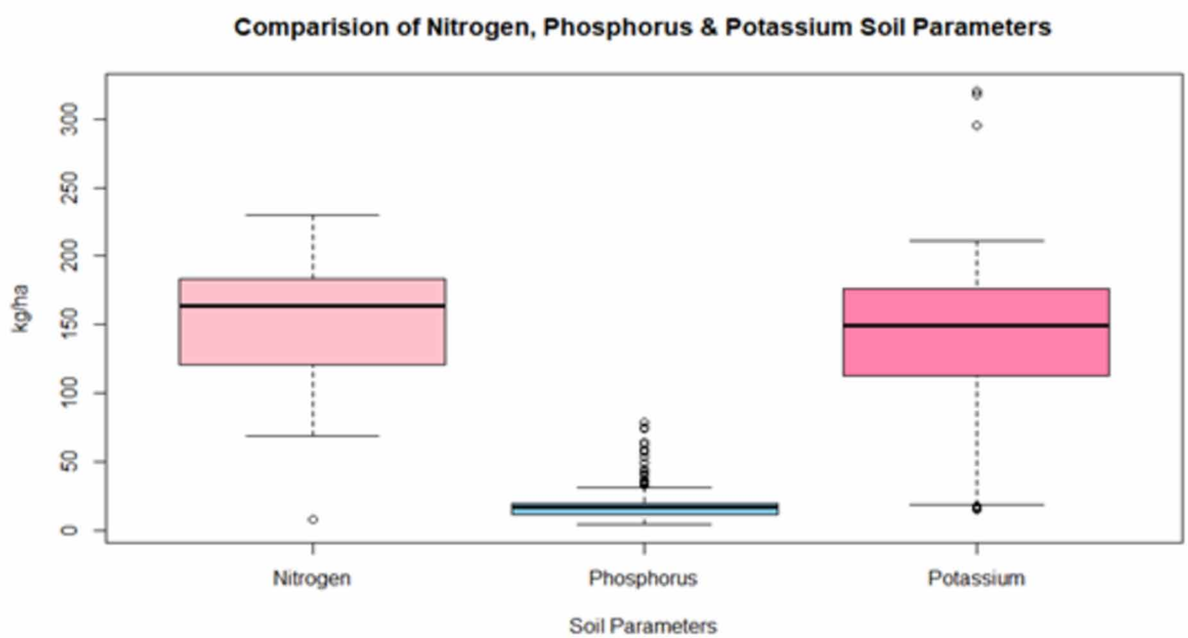


Figure 9. Comparison Box-plot for nitrogen, potassium and phosphorus



Problem-Formulation

Consider a set of spatial-data points $P = \{ P1, P_{2,} P_{3,} ....P_n\}$in space associated with the dimension *(ρ≥1)*, an attribute-function $f^{attr}()$ is defined by mapping P to R (set of real numbers). The attribute-function $f^{attr}(P_i)$ refers to the attribute-value of spatial point $P_{i.}$ For such given spatial- point 'P_i', let $NN_k(P_i)$ denotes the k-Nearest-Neighbor of point P_i, where $k = k\ (P_i)$depends on 'P_i' values with $i = \{1, 2, 3, ..., n\}$.

A neighborhood-function $f^{aggr}()$ is defined by mapping P to R such that, for each (P_i), function $f^{aggr}(P_i)$ returns the statistics summary of the attribute-values associated with k- Nearest Neighbour(P_i). To detect/remove spatial-outliers, authors have compared the attribute– values of every point (P_i)with the neighboring attribute-values $NN_k(P_i)$. Here, the comparison is carried by a comparing function $F^{ratio}()$ which refers to the function of $f^{attr}()$ and $f^{aggr}()$.

Further, Let $r_i = F^{ratio}(P_i)$ for $i = \{1, 2, 3... n\}$. For the given attribute-function $f^{attr}()$, function $k()$, neighbor function $f^{aggr}()$ and a comparison function $F^{ratio}()$ .A point 'P_i' is identified as spatial-outlier or S-outlier, if 'r_i' represents an extreme-value of the given set ($r_1, r_2, ..., r_n$). Here, it is noted by the definition that, it mainly depends on the function choices of $k(), f^{aggr}(), F^{ratio}()$ which represents more generalized.

Algorithm Defined: Iterative-R Algorithm

For simplicity, the description makes an assumption that all $k(P_i)$ are equal to a defined fixed number 'k'. But the algorithm can be made more generalized by replacing the fixed k-value with dynamic k-value $k(P_i)$. The neighboring-function $f^{aggr}()$, evaluated for a spatial-point 'P_i' is considered to be the average-attribute value of all the k-Nearest Neighbor of 'P_i'. The comparison function $F^{ratio}(P_i)$ is considered to be the ratio of $f^{attr}()/f^{aggr}()$. Very large/small values of $F^{ratio}(P_i)$ (detected by threshold-Θ) gives an indication that 'P_i' may be an S-outlier.

Input:

- Spatial-Dataset with coordinate's sample-values for Davangere region.
- Attribute-function: $f^{attr}()$
- Number *"k"* for Nearest -Neighbor and an expected number *"n"* for spatial-outliers.

Output:

- Spatial-outlier detection/removal based on the referred-coordinated sample values with crop-coordinates.

Algorithm:

Step 1: For each spatial-data point 'P_i', compute the k-Nearest Neighbour set $NN_k(P_i)$, the neighborhood-function

$$f^{aggr}\left(P_i\right) = \frac{1}{k} \sum_{P \in NN_k(P_i)} f^{attr}\left(P\right)$$

and the comparison-function $F^{ratio}(P)$, $r_i = F^{ratio}(P_i) = f^{attr}(P_i)/f^{aggr}(P_i)$.

Step 2: Let 'r_q' and 'r_q^{-1}' denote the max/min-value of 'r_1', 'r_2', ..., 'r_q', 'r_1^{-1}', 'r_2^{-1}', ..., 'r_q^{-1}'. For the given threshold (Θ), if 'r_q' or 'r_q^{-1}' $\geq$ 'Θ', Consider the respective 'P_q' as S–outlier.

Step 3: Update $f^{attr}(P_q)$ to $f^{aggr}(P_q)$, then for every spatial-point 'P_q' update $f^{aggr}(P_i)$, and r_i.

Step 4: Repeat step 2 and step 3 until threshold-condition is not satisfied or the total number of S-Outliers become equal to 'n'.

The outlier free data is further applied to Management-Zone-Delineation to improve the yield productivity. Upon removing the outliers the data still exist in the undesirable format. Thus log transformation methods are applied to make the data in the desired format. The main idea to use log-transformation technique is the data-skewness, where one/few points are much larger than the bulk data. During data pre-processing it is explored that data is skewed towards right and left for phosphorus and potassium data-properties respectively.

At the end stage, log-transformation is made more skewed distributions to less skewed, thus making skewness towards larger values. This significant for both generating the data-patterns more interpretable and to meet the assumptions of inferential statistics.

Implementation Details of Management Zone Delineation

Problem Formulation

Currently, Precision Agriculture supports huge collection of data via sensors with high resolution that usually reflects the heterogeneity of any given natural-field. Here, fertilization plays a major role in precision agriculture which describes the availability of minerals like N, P, K, pH for crops. As, majority of the farm-fields are heterogeneous the challenging task is to identify the sub-fields and the amount of fertilizers required. This task leads to concept called as Management-Zone-Delineation. From the literature review, it is explored that the delineated management-zones are tailored to a particular application where each approach is inherently-subjective and no generic approach fits the application. Thus, farmer needs to enhance the understanding of the farm area and get the insights on the availability of minerals. Identifying the zones using the common methods like Principal-Component-Analysis, linear-combinations of the data or data-transformation leads to crisps-zones and straight computations (Horvath et al., 2012). In such scenario, understanding the newly generated variables is lost. The major issue with spatial-dataset is the adoption of non-spatial methods like k-means/fuzzy c-means that leads in zone-discontinuity which is smooth/fixed in few cases and finally present unstable-results. This problem aims at applying MZD technique to process the raw dataset for zone delineation.

Automatic Dataset Interpolation

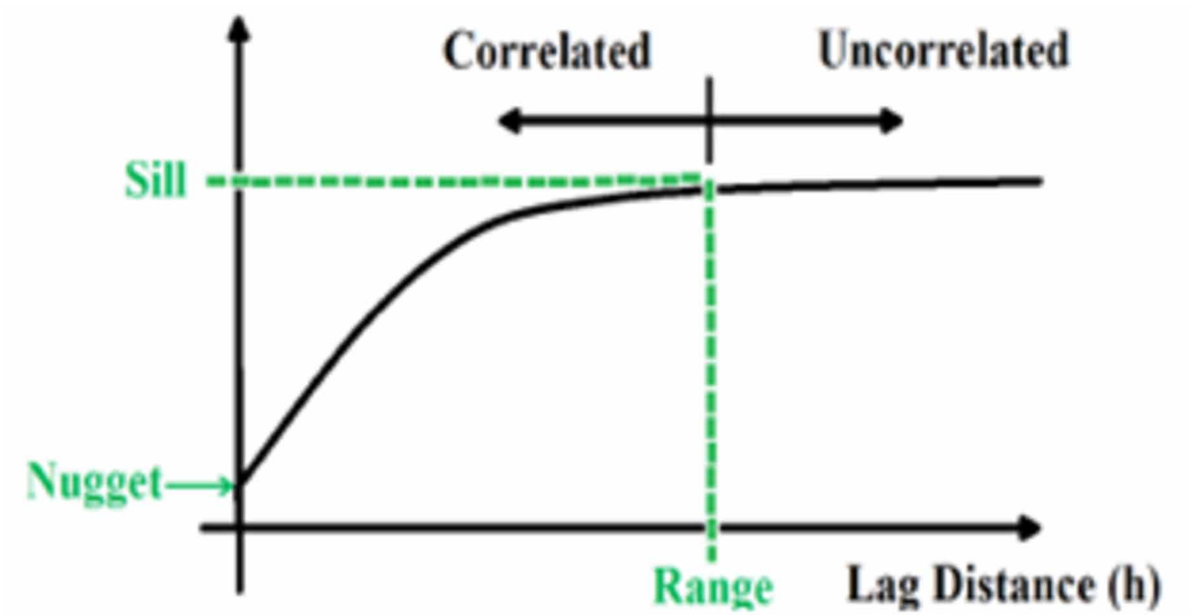
Currently, various techniques exist to interpolate spatial-data like kriging, indicator-kriging, universal-kriging, co-kriging and many more. The choice of the techniques is mainly depends on the data-characteristics and the type of spatial-model adopted. The most commonly adopted method is ordinary-kriging and is selected in this research work.

Kriging is an efficient choice for an estimator, as the work considers the statistical-data. The standard co-variance models such as variogram are employed for data estimation that are same for every-year and spatial-anisotropy can be exhibited easily. The primary reason to select kriging technique is to maximize the variance error which is nothing but the kriging-variance. This mainly depends on the data configuration and the variance which is regarded as homoescedasic in nature. Interpolation-variance is the precise measure of accuracy for estimating the spatial-data (particularly for skewed-data).

Mathematical Model: Kriging and Semi-Variogram

Interpolation technique (kriging) is used to achieve Management-Zone-Delineation which is segment of geo-statistics that is adopted for surface-mapping from minimum-sample data and efficiently used to estimate the values of un-sampled locations. It also determines the geological characteristics of the areas between the measured data points in geo-statistical applications (Celal Guvendik et al., 2012). Here, kriging utilizes the weighted-average of the neighboring sample-points to determine the unknown location values from the given known-locations. Thus, the weighted average of the parameter values of

Figure 10. Semi-variogram model



commonly used models can be a better way of describing the spatial processes (AsimBiswas etal., 2013) Semi-variogram model is used to optimize the weight which is shown in Figure 10.

Spatial soil variations are quantified by employing the semi-variance which is calculated as the squared difference among the variable values at two separate locations.

$$\gamma(h) = \frac{1}{2} E\{[Z(x) - Z(x+h]^2\} \tag{1}$$

where, *(Z(x))* = Refers a variable-value of interest *(soil-parameters N, P, pH, K)* for the given Location *{x,y}*. *E{ }* = Refers to the statistical expectancy operator

From Figure 11, it can be presented that semi-variance is concerned of how large the spatial-variance as a function of distance among two-locations. *E{ },* in Equ(1) refers to the expected-value which inferences that larger the distance among two-locations greater will be the difference among the soil-properties. As, there is increase in the distances the correlation decreases and turns to zero at some-points. The point where correlation becomes zero is termed to be the “Range of semi-variogram” and indicates the distance upto which there exists the spatial-correlation.

Sill refers to the maximum-value of semi-variogram which is the dataset-variance, Nugget refers to the measurement-error. The major challenge is to decide the precise predicted-value for every uniform location when only data associated to spatial-location is available.

Thus, to simplify the above problem, the authors have refined the statistical expectancy operator in Equation (1) by restoring *‘x’* by *‘u’* and then the updated estimator (*Z*(u)*) is obtained as given below;

$$E\left\{\left[Z'(u) - Z(u)\right]\right\} = 0 \tag{2}$$

Figure 11. Distance among two-locations

Consider the scenario when the estimation is needed for only single-location. Then, the simple linear-sum is defines as;

$$Z'(u) = \sum_{a=1}^{n} \lambda_a Z\left(Z\left(u_a\right)\right) \tag{3}$$

Substituting Equation (1) in (2);

$$\sum_{a=1}^{n} \lambda_a E\left[\mathrm{Z}\left(u_a\right)\right] = \sum_{a=1}^{n} \lambda_a E\left[\mathrm{Z}\left(u\right)\right] \tag{4}$$

Suppose, the expected-value is static (constant over domain), then unbiased-condition becomes;
Substituting Equation (5) in (4):

$$\sum_{a=1}^{n} \lambda_a = 1 \tag{5}$$

$$E\left[Z\left(u_a\right)\right] = E\left[Z\left(u\right)\right] \tag{6}$$

The assumption made for the static expected-value is occasionally used in practicable modeling and can be decomposed into mathematical units as;

$$Z(u) = M + R(u) \tag{7}$$

where, M= Refers to some constant unknown/expected value of mean(m) and Range(u) and represents the residual term that is varies spatially. Thus Equation (7) changes to:

$$Z(u) = m + R(u) \tag{8}$$

By, expanding the above equation:

$$Z(u) = m(u) + R(u) \ \text{with} \ E\left[R(u)\right] = 0 \forall u \tag{9}$$

This means,

$$E\left[Z(u)\right] = E\left[m(u)\right] \tag{10}$$

By, substituting the mean-value on both sides in Equation (2):

$$E\left\{\left[Z^{'(u)}\right] - m(u)\right\} = E\left[Z(u) - m(u)\right] \tag{11}$$

Modifying Equation (11) results in;

$$E\left\{\left[Z'(u)\right] - m(u)\right\} = \sum_{a=1}^{n} \lambda_a E[Z(u_a) - m(u_a)] \tag{12}$$

Further, when estimator is applied for only one-location then Equation (12) becomes;

$$Z'(u) - m(u) = \sum_{a=1}^{n} \lambda_a [Z(u_a) - m(u_a)] \tag{13}$$

Finally, the Equation (13) obtained refers to the "kriging Equation". Where,

- **u_α, u:** Refers to the location-vectors for the estimated point. Here, the neighboring data-point is termed by **α**.
- **(n(u)):** Refers to the total number of data-points in the local neighborhood used to estimate (**Z*(u)**).
- **$m(u_\alpha)$,m(u):** Refers to the expected mean-values for **Z(uα)** and **Z(u)**.
- **$(Z(u_\alpha))$:** Refers to thekriging-weight assigned to each node.
- **(Z(u)):** Refers to random-field with m(u) (mean) and R(u) (residual-component)

R(u) = (Z(u)) - m(u)

Model Adopted: Semi-variogram
M. Stein's and Ste-Matern Parameterization:

$$\gamma\left(h\right)=c_0\left[1-\frac{2}{\gamma\left(v\right)}\left(\frac{h\sqrt{v}}{\alpha}\right)^{v}K_v\left(2\frac{h\sqrt{v}}{\alpha}\right)f\ or\ h>\pm\right] \tag{14}$$

where, $v>0$:Refersto the smoothness-factor parameter. Matern-Model:Refers to the special exponential-case and Gaussian-Model. Matern semi-variance: Refers to semi-variance class which provides flexibility for describing the spatial-autocorrelation.

Description of the Work Implemented: Management-Zone-Delineation

Management-Zone-Delineation is accomplished by an interpolation method known as kriging, which predicts the function value at any given data-point by calculating the weighted-average value of the known sample-values in the neighborhood location-point. Here, estimation is carried out to realize the random-fields. Adopting kriging, a methodology foundation can be built to the spatial-influence for the unobserved-locations and also quantify the unpredictable related to the estimator.

Kriging allocates weight in-line with the data-driven weighing-function, which is defined with certain standard rules:

1. A better result can be estimated, if the data-locations are fairly-dense and is uniformly scattered over the grid or the map-area.
2. An unrealizable- result is estimated, if the data-locations fall in few-clusters with huge-gaps.

To Achieve Management-Zone-Delineation, Two Important Tasks Are Required

Firstly, the task should exhibit the set of dependency rules. Secondly, to build the predict-models. To accomplish these two tasks, kriging involves two-step process. Initially, it creates variogram and few covariance-functions for estimating the spatial-autocorrelation that mainly depends on the model adopted. Further, it extends to predict the unknown values. The major soil-sample properties are quantified and their respective surface-maps are prepared. Finally, it involves in analyzing the gathered information from the soil-map, which helps t make the decision on applying the percentage of fertilizers/pesticides with regard to the amount of N, K, P and pH for the given soil-sample.

Efficient interpolation and analysis of the soil-map leads to decision-support-system for variable-soil treatment within sub-fields. This approach aims at providing a precise-resource application which optimizes the farmer energy, time, economic-benefits, crop yield-productivity which together results in precision-agriculture. However, since the fields are usually heterogeneous, different parts of the field may require different amounts of basic fertilization (George Rub et al., 2011).The delineation of MZ affords the opportunity of variable rate application of fertilizers and the minimization of pollution risk due to an excessive application of resources (NahuelRaúl Peralta et al., 2015).

Pseudo-Code for Management-Zone-Delineation (MZD)

The presented algorithm for MZD depends on interpolation method called kriging and k-means clustering algorithm.

```
for each soil-sample point in the given spatial-dataset
   {
convert spatial-coordinates into spatial-object
for each spatial-object created
         {
   locate spatial-object on Earth's surface using Coordinate-
Reference-System(CRS)
   convert the coordinates of the input-data into WGS84-CRS
      call function MZD( input-data points)
          } }
function MZD( input-data points, formula)
     {
   for the given heterogeneous farm-field
   call function krig(new-data points, formula)
      }
functionkrig( new-data points, formula)
     {
   delineate the heterogeneous field into homogenous sub-fields
returns the precise percentage of fertilizers required for each
sub-zones
   }
```

Results and Analysis

Analysis on the Results Obtained for Outlier Detection/Removal

The survey carried out for soil physico-chemical parameters has the following ranges: Nitrogen - (48 to 236 kg/ha), Phosphorus - (8 to 78 kg/ha), Potassium- (9 to 312 kg/ha) But, from the obtained dataset it can be summarized that 'n' ranges from 7.08 to 230.10 kg/ha, 'p' ranges from 4.32 to 78.90 kg/ha and 'k' ranges from 5.06 to 319.56 kg/ha. When checked for conformity of each of n, p, k values of the prepared dataset with those values from the survey standards, it is observed that out of 652 observations 23 non-conforming values for 'n', 77 non-conforming values for 'P' and 29 non-conforming values for 'k' attained is shown in Figure 12, Figure 13 and Figure 14.

Graphical spatial-outlier is performed which relies on the spatial-data visualization to discover the spatial-outliers and variogram gives the description of the spatial-data continuity. The variogram-plot is applied as a fitted-model to check the correlation among the spatial-observations. Whereas, semi-variance depicts the spatial auto-correlation of the measured data-points. The distance in which the model initially flattens-out is called as range. The locations closer to the range-value are spatially auto-correlated and farther locations than the range are not. The value at which semi-variogram attains range-value is re-

Figure 12. 629 complying values of 'N'

```
[1] "clean observations for n:"
  [1] 103.210 124.020 102.310 160.160 162.010 153.020 102.320 143.020
  [9] 102.320 102.310 173.020 184.760 178.750 194.070 112.750  75.060
 [17]  50.015  54.070 112.320 109.410  70.520  52.101 108.310  97.010
 [25]  68.010  49.020  58.150 109.150  97.320 112.310 197.140  48.320
 [33] 154.010 194.760 198.070 193.010 114.500  68.010  73.010  93.010
 [41] 112.010  68.520 102.350  97.320 157.150  76.160 125.020  64.750
 [49] 102.310  68.150  53.010 103.250  53.020  64.070 102.320 107.100
 [57]  84.010 102.150 102.750 110.730 184.320 184.150 198.100 170.650
 [65] 159.320 148.310  78.170  98.150  98.010 125.020 107.910  97.070
 [73]  73.020 117.150  97.020 124.010  68.320 191.010  98.160 106.310
 [81] 114.750  74.010  48.170  97.410  65.010 102.110 106.750 120.750
 [89] 108.320 112.010 114.150 193.010 168.750 186.710 102.320  65.150
 [97] 193.010 112.150  90.310  63.020 173.100 173.010 197.020 168.010
[105] 183.020  83.020 165.060 197.040 163.070 194.010  94.320  97.010
[113]  83.020  97.040  88.020  63.010 104.460  70.160  98.150  84.010
[121]  59.160  99.670  70.250 112.320  63.020  63.070 102.310  50.410
[129] 102.750 106.160 102.310 107.040  64.050  98.150  60.010 102.310
[137] 124.020 114.310  58.320  62.310 108.150 105.710  98.150  74.000
[145] 118.070  98.020  48.750  67.150 112.320  68.150  64.150  73.020
```

Figure 13. 575 complying values of 'P'

```
[1] "clean observations for P:"
  [1] 22.31 27.08 12.05 23.01  8.15  9.75 19.75 14.01 16.01  9.75 20.75  9.32
 [13]  8.32 13.27  8.32  8.01  8.01 12.65 11.75 13.75 12.54 20.15 11.01  8.55
 [25] 24.15  9.32 13.15 12.06 14.06  8.76 12.75 20.65 22.15 22.75 12.17  8.32
 [37]  8.11  8.31 16.75 16.32 14.76 22.31 20.01 10.11 10.15  9.15 13.81 16.16
 [49] 22.31 12.75  9.32 10.15  9.15 20.16 16.75 16.35  8.75 18.15  9.34 12.15
 [61]  8.67 12.15  9.17  9.02 17.15  8.32 24.15 23.01  9.01 13.15 12.75 20.32
 [73] 12.75 17.81 17.02 17.15 12.65 22.31 18.76 17.32 32.01  9.32 20.15 24.15
 [85]  8.32 24.01 16.05  8.15 15.07  8.32  9.16 17.02  8.75  8.31  8.31 14.15
 [97]  9.75 16.45  9.32 20.15 16.31  9.01 53.52 16.75  8.15 14.01 13.25  8.26
[109] 10.71 14.20  8.15 16.75 14.15 12.01 20.16  8.35 16.32 18.15 40.25 22.31
[121] 14.01 16.15 10.32 16.75 18.75 18.73  9.32 15.32 17.32 20.06 18.75 13.27
[133] 16.72 24.32 16.01  8.07 14.32  9.14 16.54 12.61 34.15  8.31  9.06  9.32
[145]  8.77 16.31  9.32  8.15 22.15 12.75  9.32 12.75  9.75  9.73 14.75 13.75
[157] 15.60 21.41  8.32 14.32 16.15 56.60 16.32 12.15  8.15 16.15 24.15 24.02
[169] 16.75 22.15 18.32 10.15 18.36  9.04 18.75  8.79  8.75 24.15 57.02 62.59
[181] 20.53 24.16  9.75  9.05 12.32 18.75 17.72 16.32 15.32 22.15 21.57 10.12
[193] 18.75  9.75  8.32 22.10 14.05 18.75 14.75  9.32 16.75  8.15 16.75 12.75
[205] 16.75  9.01 12.16  8.14 10.32 24.75  9.15 22.01 24.01 15.02 12.02  8.75
[217] 12.06 35.20 10.71 74.11 20.08  8.14  8.32  9.31  8.16  8.06 13.07 16.14
```

garded as sill. The nugget-effect refers to the measurement-errors/spatial-sources of variance at smaller distances than sampling-interval or both. The micro-scale variations lesser than the sampling-distances remain as a part of nugget-effect.

The nugget-value for nitrogen (N) is 343, which specifies that outlying-observations exists due to the measurement-errors. Whereas, the sill-value is 2298, beyond this value there exists no correlation and indicates that the variable is purely-random and thus the variogram flattens. The range-value is 11361 at which the variogram-model initially flattens.

Figure 14. 623 complying values of 'K'

```
[1] "clean observations for K:"
  [1] 105.32 124.63  58.80  58.22  58.32  69.73  68.11  65.01  78.02  83.02
 [11]  53.02  97.97  89.76  50.67  68.75  49.75  82.07  82.07 102.31  90.01
 [21]  80.32  78.11 124.02 102.31  59.04  54.05 112.75  92.15  43.07 102.73
 [31]  59.17 109.31  73.16  49.76  63.14  86.04 121.26  90.49  93.01  43.01
 [41]  59.06  65.01  74.32  97.32  68.71  53.02  61.75  75.06 123.01  43.20
 [51]  49.01  68.07  70.15  97.32  64.15  64.05  58.15  43.65  48.10  93.16
 [61]  94.23 125.32  35.65 104.15  94.15  53.02  97.01  63.07  78.15  98.09
 [71]  54.01 109.31  61.06  72.15  98.17  90.15  53.16 129.15  59.75  62.72
 [81]  98.15 102.75  97.08 102.31  68.15  63.02  75.02  93.14  67.74  93.65
 [91]  74.75  90.36 123.01  68.01  79.06  70.07  79.02  53.15  68.01  94.01
[101]  53.02  93.15 102.16  89.32  76.02  34.01  37.02  37.02  79.06 117.65
[111] 102.32  71.01 124.75 102.31  97.66  64.01  65.56  45.75 107.65  78.06
[121] 103.06  96.75  54.75 102.15  80.16  93.02  93.17  83.01  84.05  78.72
[131]  60.14  97.02  90.11  37.96  91.43  82.13  83.24  64.07 102.31  78.15
[141]  94.07  79.07  83.01  83.01 102.77  45.04  59.12  74.41  93.16  32.07
[151]  93.02 294.24  90.71  38.15  70.78  48.06 117.05  49.15  93.15 103.31
[161]  54.15  68.15  70.16  60.02  43.02  94.01  63.21  97.05  48.15  59.32
[171]  94.75  80.67  94.01 103.65  59.01  73.01  93.01  97.63  83.12  73.15
[181]  68.75  24.07  70.16 117.48  98.16 118.75  80.10 112.31 125.01  93.01
```

The nugget-value for phosphorus (P) is 113, which specifies that outlying-observations exists due to the measurement-errors. Whereas, the sill-value is 150, beyond this value there exists no correlation and indicates that the variable is purely-random and thus the variogram flattens. The range-value is 253972 at which the variogram-model initially flattens.

The nugget-value for potassium (k) is 103, which specifies that outlying-observations exists due to the measurement-errors. Whereas, the sill-value is 2169, beyond this value there exists no correlation and indicates that the variable is purely-random and thus the variogram flattens. The range-value is 19026 at which the variogram initially flattens.

The details of the above discussed observation are shown in Figure 15.

The plotted graph shown in Figure 16 represents N, K, P values for 685 locations of different villages from six-talukas of Davangere-District using scatter-plot.

To perform the detection/removal of outlier-observations, initially KNN-algorithm is applied to find the Nearest-Neighboring-points. It is observed that large data-points exist in the low-dimensional space. For KNN-Classification, the input-data consists of k-closer training samples in the feature-space and the output-data consists of class-membership.

The Euclidean-Distance measure is used to examine the set of k-instances in the given training-dataset that are most similar to the new-input. The value of *'k'* is set to $k=23$, as a standard rule-of-thumb to choose the k-value is given by $k = \sqrt{N}$ where, *'N'* refers to the number of sample-points in the training-data (685-observations).The nearest-neighboring points is located using KNN-algorithm and is shown in Figure 17.

Further, an attempt is made to locate the outlier by applying Iterative-R (Ratio) algorithm that performs for multi-iterations, where it identifies one-outlier in every iteration. So, that the identified outlier does not have impact on the subsequent-iterations negatively. To illustrate the working procedure of the adopted-algorithm, the same dataset is considered with the initial 24-observations only which is shown in Figure 18.

Figure 15. Semi-variogram Model for (a) Nitrogen (b) Phosphorous (c) Potassium

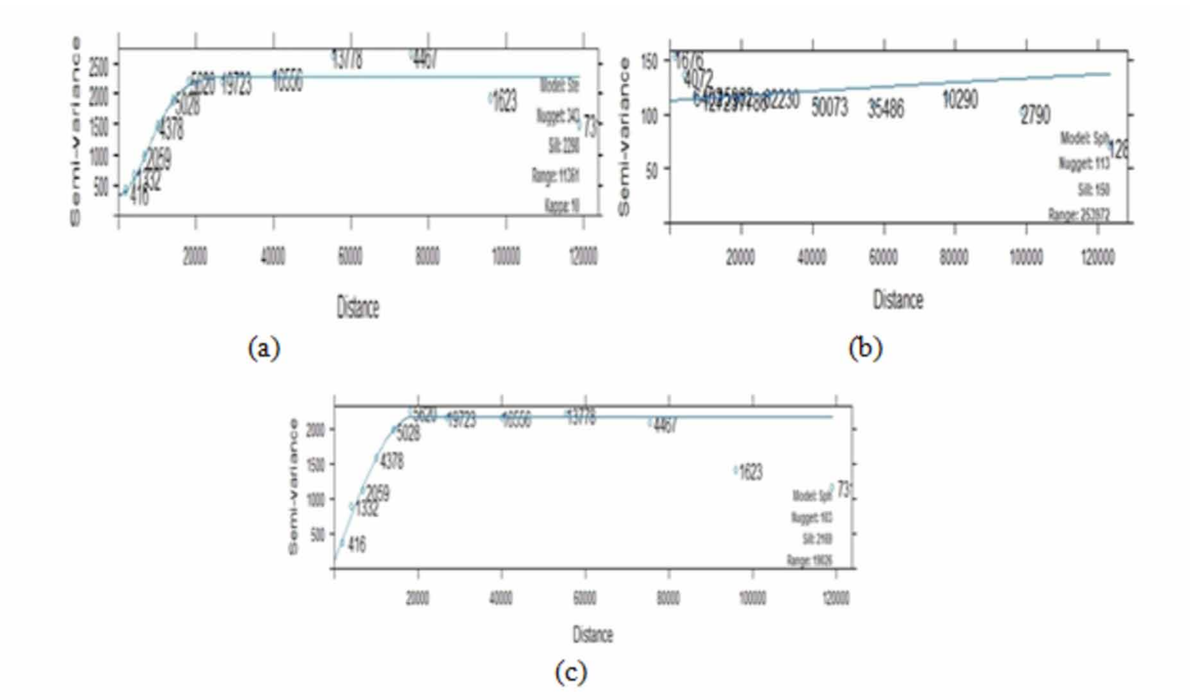


Figure 16. N, K, P values for given sample-Dataset

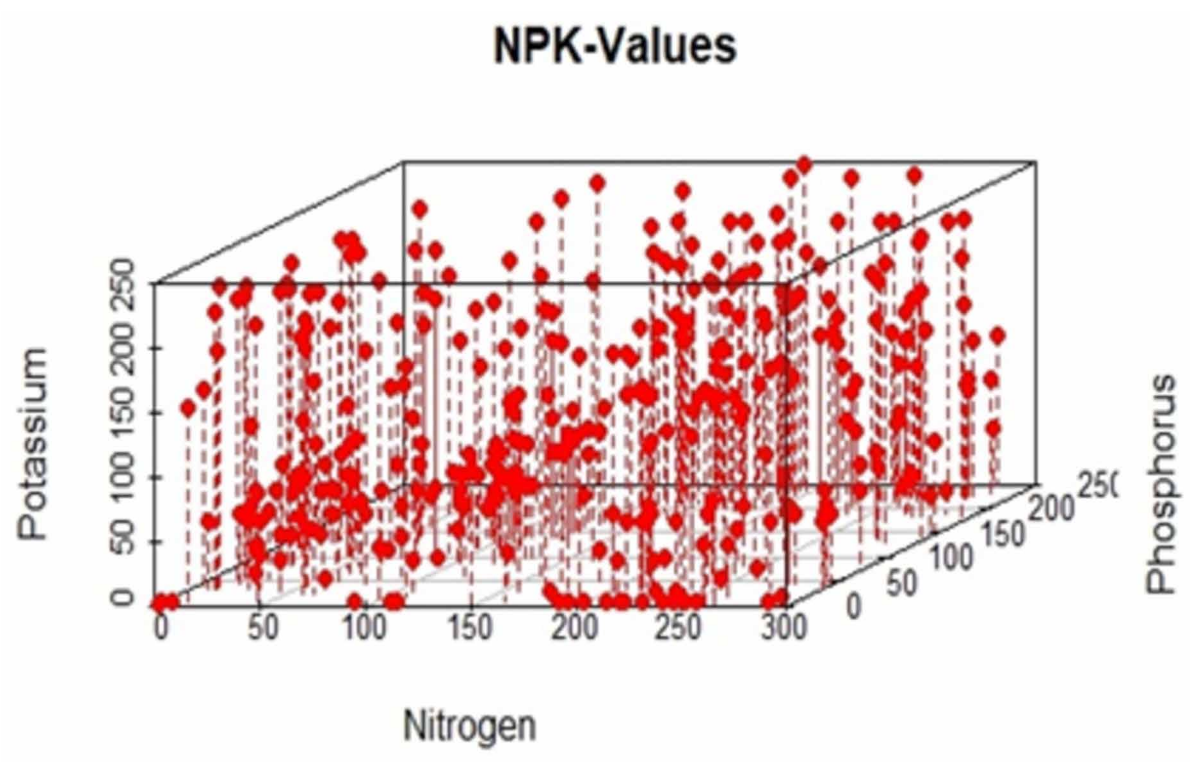


Figure 17. Nearest-Neighbors points located by KNN Algorithm

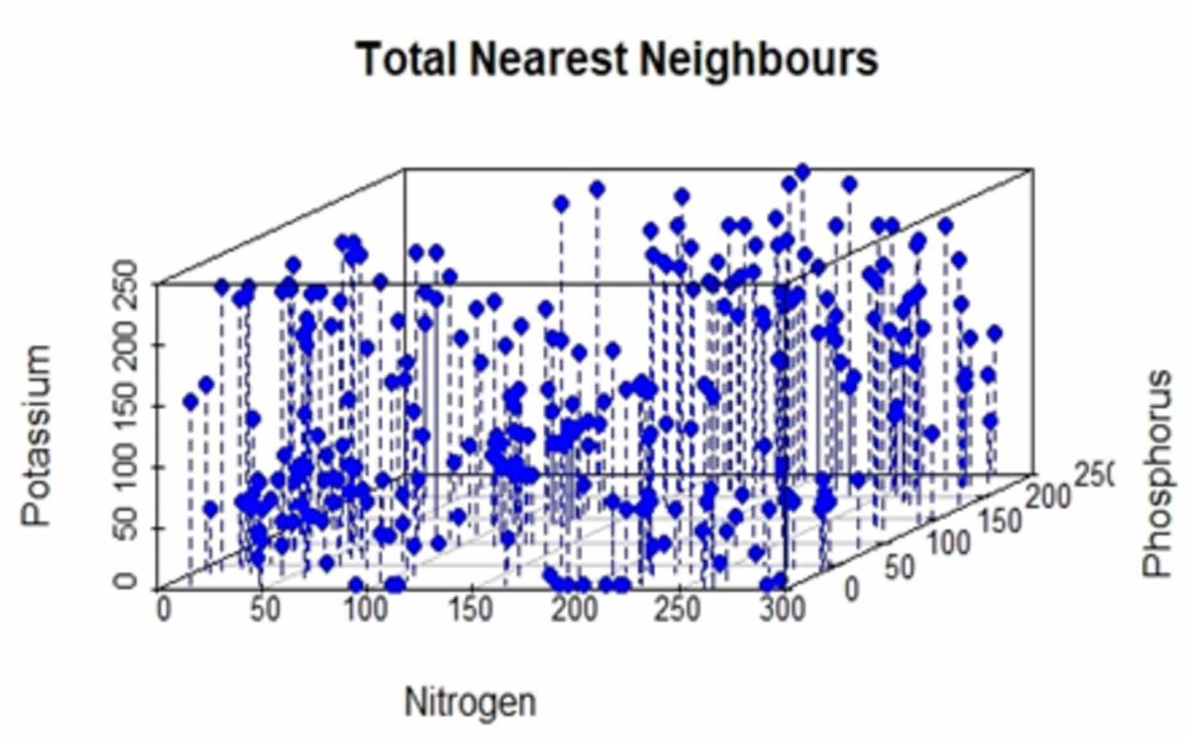
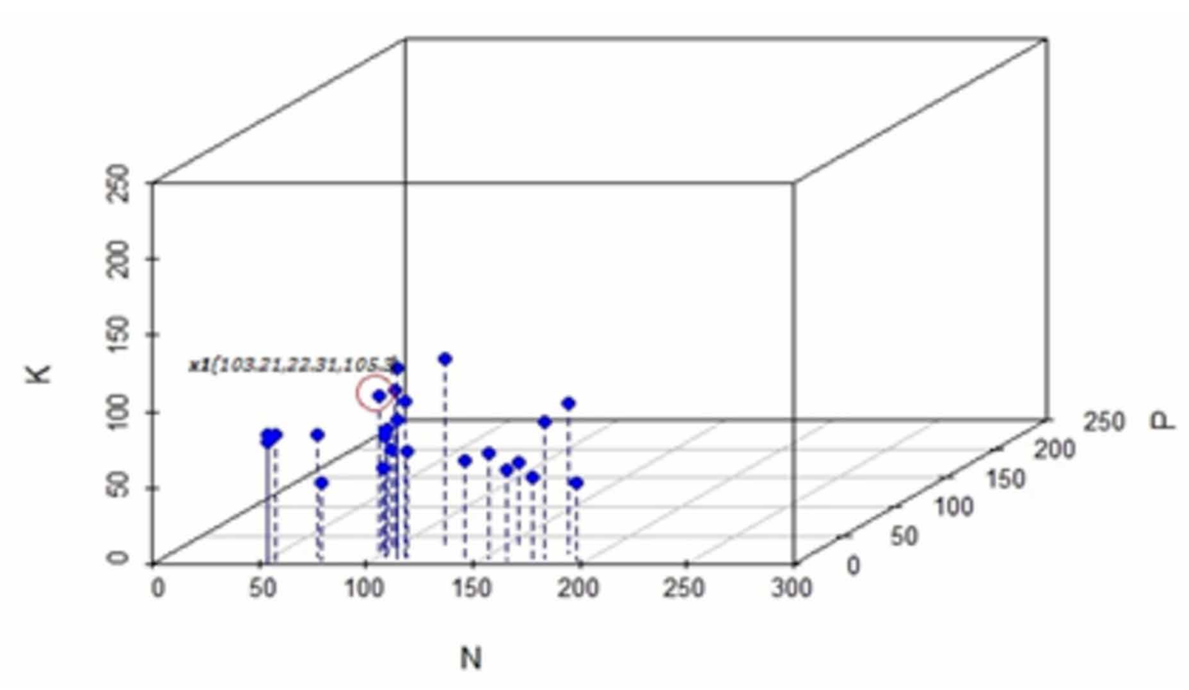


Figure 18. Initial 24-observations



The corresponding scatter-plots for the initial 24-observations for the given soil-parameters N, K, P values is shown in Figure 19 and Figure 20.

Initially, compute the distance between each single-data point (P_i) and remaining data-points using Euclidean-Distance-Formulas as given in Figure 21.

Next, the value of k-input is set to be $k=10$, to retain only those $NN_k(P_i)$-points that are more closer to any (P_i)- point than the rest of the points and then retrieve positions of the nearest-neighbors which is shown in Figure 22.

Figure 19. Scatter-Plots for initial 24-observations for (a) 'N' (b) 'P'

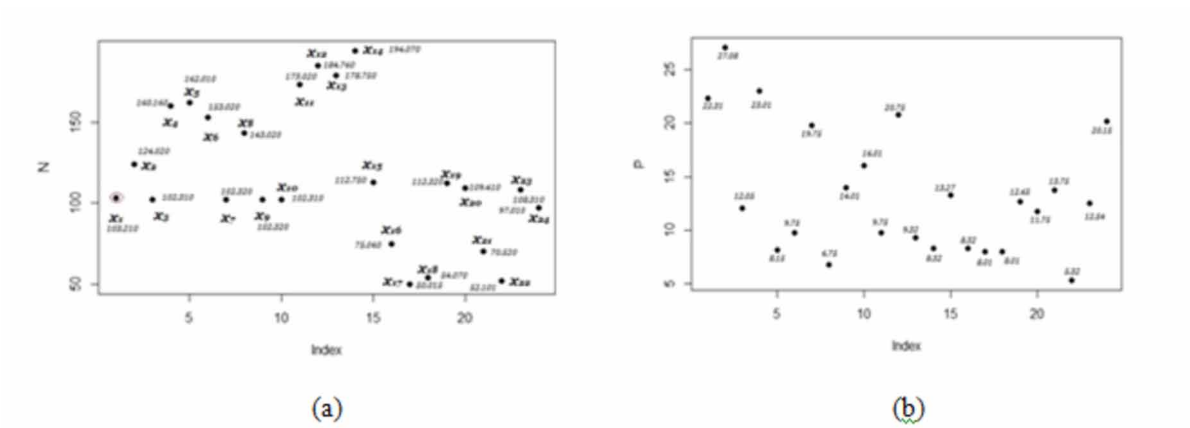


Figure 20. Scatter-Plots for initial 24-observations for 'K'

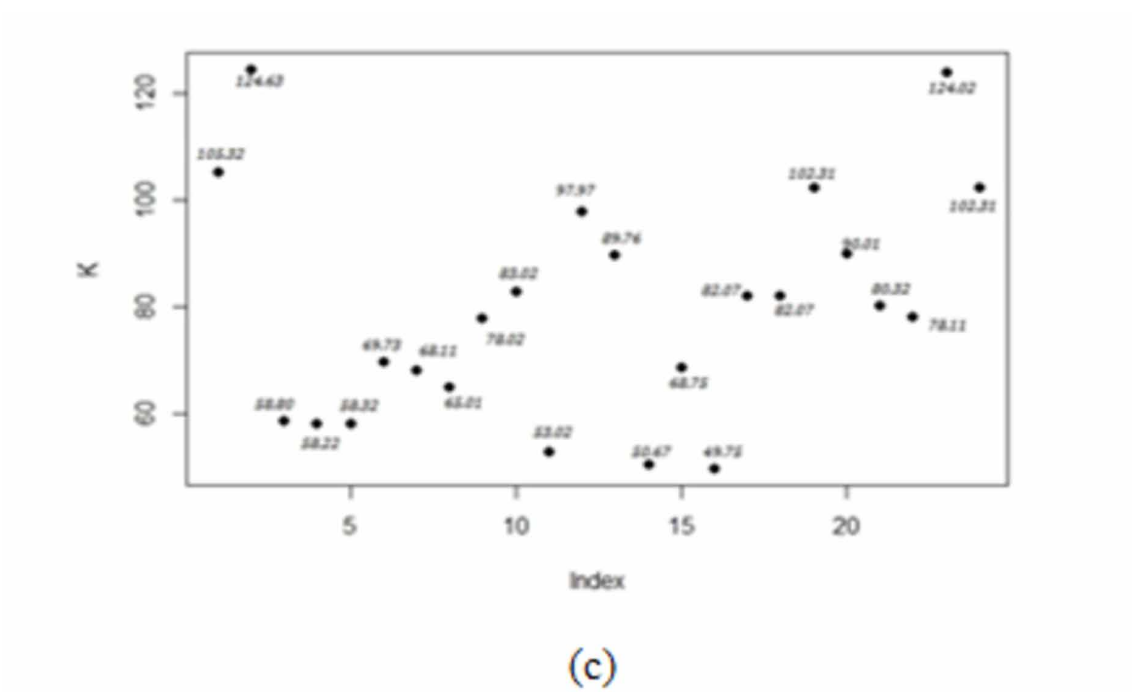


Figure 21. Euclidean-Distances computed (each single data-point (P_i) and remaining data-points

```
 [1]  0.000 44.890 57.680 10.550  2.360  1.660 40.660 16.060 36.490 29.500  4.950 72.640 46.990
[14] 22.220 36.070 97.710 90.745 86.690  3.560 19.670 66.250 95.309 14.030 11.370

 [1]  44.890   0.000 102.570  34.340  47.250  43.230  85.550  60.950  81.380  74.390  39.940
[12]  27.750   2.100  22.670  80.960 142.600 135.635 131.580  48.450  64.560 111.140 140.199
[23]  30.860  56.260

 [1]  57.680 102.570   0.000  68.230  55.320  59.340  17.020  41.620  21.190  28.180  62.630
[12] 130.320 104.670  79.900  21.610  40.030  33.065  29.010  54.120  38.010   8.570  37.629
[23]  71.710  46.310

 [1]  10.550  34.340  68.230   0.000  12.910   8.890  51.210  26.610  47.040  40.050   5.600
[12]  62.090  36.440  11.670  46.620 108.260 101.295  97.240  14.110  30.220  76.800 105.859
[23]   3.480  21.920

 [1]  2.360 47.250 55.320 12.910  0.000  4.020 38.300 13.700 34.130 27.140  7.310 75.000 49.350
[14] 24.580 33.710 95.350 88.385 84.330  1.200 17.310 63.890 92.949 16.390  9.010

 [1]  1.660 43.230 59.340  8.890  4.020  0.000 42.320 17.720 38.150 31.160  3.290 70.980 45.330
[14] 20.560 37.730 99.370 92.405 88.350  5.220 21.330 67.910 96.969 12.370 13.030

 [1]  40.660  85.550  17.020  51.210  38.300  42.320   0.000  24.600   4.170  11.160  45.610
[12] 113.300  87.650  62.880   4.590  57.050  50.085  46.030  37.100  20.990  25.590  54.649
[23]  54.690  29.290

 [1] 16.060 60.950 41.620 26.610 13.700 17.720 24.600  0.000 20.430 13.440 21.010 88.700 63.050
[14] 38.280 20.010 81.650 74.685 70.630 12.500  3.610 50.190 79.249 30.090  4.690

 [1]  36.490  81.380  21.190  47.040  34.130  38.150   4.170  20.430   0.000   6.990  41.440
[12] 109.130  83.480  58.710   0.420  61.220  54.255  50.200  32.930  16.820  29.760  58.819
[23]  50.520  25.120

 [1]  29.500  74.390  28.180  40.050  27.140  31.160  11.160  13.440   6.990   0.000  34.450
[12] 102.140  76.490  51.720   6.570  68.210  61.245  57.190  25.940   9.830  36.750  65.809
[23]  43.530  18.130
```

Finally, the neighbourhood-function $f^{aggr}()$ is calculated which is described as a map with P to R, such that for every (P_i), $f^{aggr}(P_i)$ returns the summarized –statistics of the attribute-value of all the given spatial-points with $NN_k(P_i)$. Consider for example, $f^{aggr}(P_i)$ can be defined as the average attribute-value of k-nearest-neighbor of (P_i) as given below:

[1}] 8.4210 29.3030 23.4284 11.5740 8.4210 8.6760 17.4510 12.2160 15.7830 14.7380

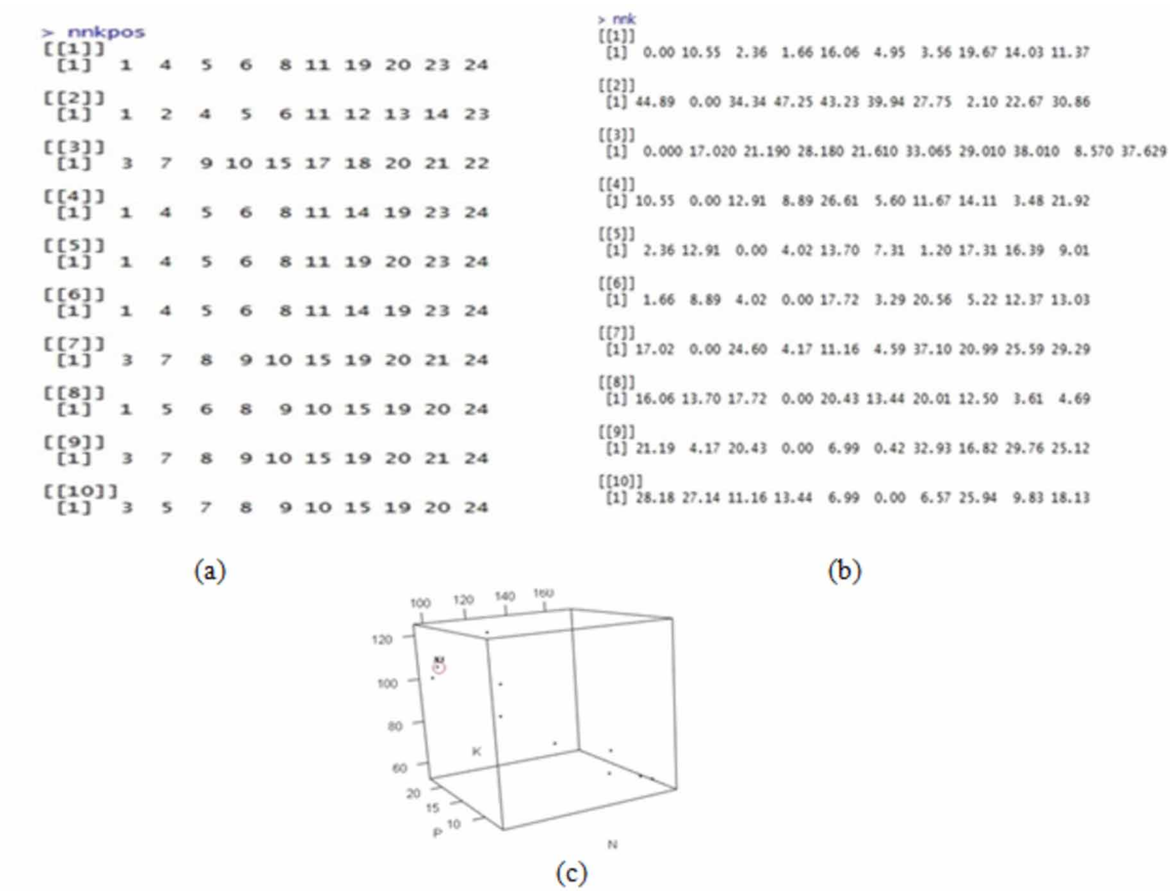
For each (P_i), its summarized attribute-value is assigned as a attribute-function $f^{attr}(P_i)$ as given below:

$f^{attr}(P_i)$ (76.94, 91.91, 57.72, 80.46, 76.16, 77.5, 63.39, 71.59, 64.78, 67.11)

The ratio (R-value) is computed for every (P_i)'s point attribute-value $f^{attr}(P_i)$ and the average attribute-value of its nearest-neighbour point $f^{aggr}(P_i)$. Then, for each-point the comparison-function $F^{ratio}(P_i)$ is defined as the ration of $f^{attr}(P_i)/ f^{aggr}(P_i)$ as given below:

[1] 9.136682 3.136539 2.463677 6.951788 9.044057 8.932688 3.632457 5.860347 4.104416
[10] 4.553535

Figure 22. Retaining $NN_k(P_i)$ data-points closer to 'P_i' (a) data-points positions (b) Distance for every (P_i) with its Closest data-points (c) 3D-Plot (Retained data-points)



Finally, the point with extreme R-value *(max(r) =9.13662 at position-1)* is declared as an outlier within the given proximity of point (P_1). Further, the attribute-value of the detected spatial-outlier is substituted with the average attribute-value of its neighbors' $f^{aggr}(P_i) = 8.421$, $f^{attr}(P_i) = 76.94$ is replaced with $f^{aggr}(P_i)$-value (8.421) as shown below:

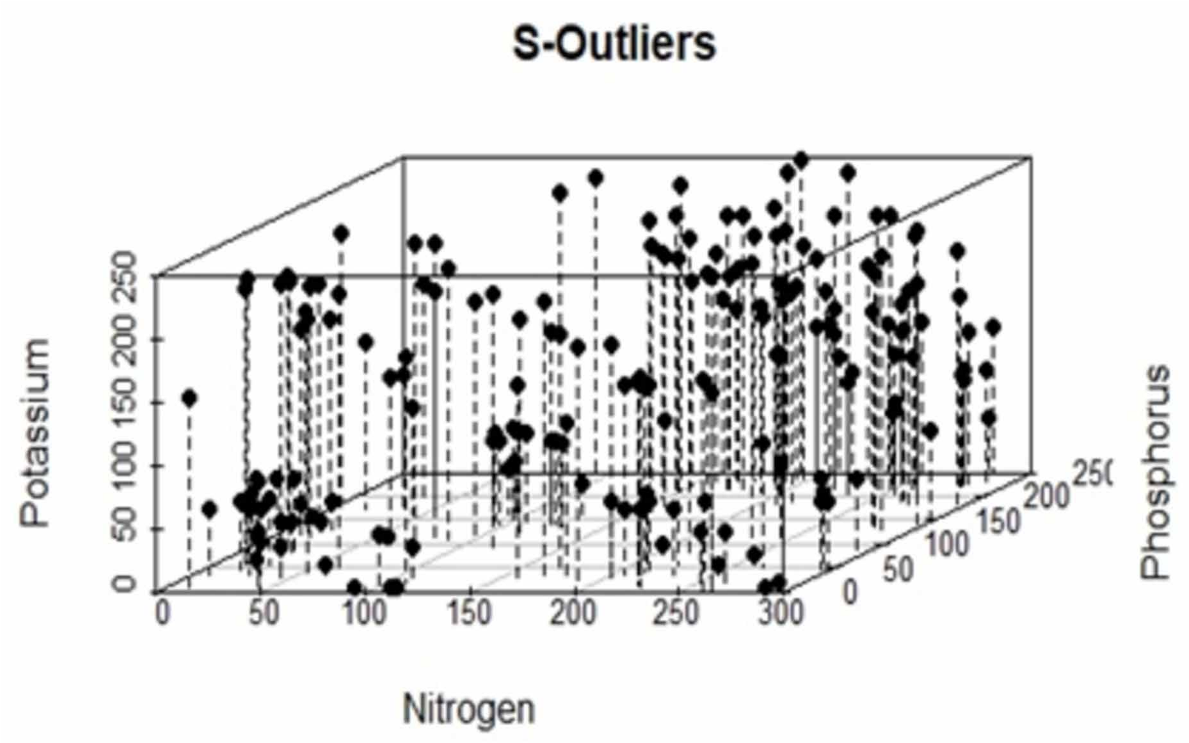
$f^{attr}(\max)f^{aggr}(\max)$

[1] 8.421 91.910 57.720 80.460 76.160 77.500 63.390 71.590 64.780 67.110

After updating $f^{attr}(P_i)$to $f^{aggr}(P_i)$, then for every spatial-point of (P_i), where $NNk(P_i)$ holds (P_q), it is required to update (r_i) and $f^{aggr}(P_i)$. The discussed steps are repeated till the threshold-criteria is not satisfied or till the total count of spatial-outliers reaches the value 'n' shown in Figure 23.

To summarize, in every iteration the ratio of the point attribute-value and its average attribute-value of its neighbor is computed for each sample-point. The point with the greater r-value is identified. The plot in Figure 23 shows the entire set of data-points that are outlier-free, anomalies generated by false-outlier identification (swamping) and neglecting few true-outliers (masking).

Figure 23. Iterative –R algorithm applied for detection/removal of outliers



Analysis on the Results Obtained for Management-Zone-Delineation

Various plots are drawn for semi-variances versus the distance among the ordered-data which is referred as semi-variogram. Semi-variogram refers to the measure of spatial-dependency between two soil-observations as a distance function. The analysis carried-out on the semi-variogram plots explains how semi-variance changes with the change in the distance among the observations.

Semi-variogram parameter refers to the characteristics of empirical-variogram which provides as the assessment for the variable-structure. The three major semi-variogram parameters are nugget-effect, sill and range.

Kriging presents a strong prediction, when applied on the input-data builds a mathematical-model with semi-variogram to create a prediction-surface and to validate it. The prediction-surface describes how precise the suggested model predicts.

Experimental Variogram and Fitted-Variogram Model for Nitrogen, Phosphorus, and Potassium

The nitrogen availability lies between 22.7 kgha^{-1} - 294 kgha^{-1}. For the given dataset of 853 observations after applying kriging on the nitrogen-value, kriging predicts (left) and the standard kriging error associated (right). Further, it divides the given farm-field domain into various-ranges like (22.7,61.46), (61.46,100.2), (100.2,139), (139,177.7), (177.7,216.5), (216.5,255.2), (255.2,294) which is marked by different colors in the map-area. Figure 24 also shows the automatically generated Variogram and it is observed that the Nugget-value is ‘0’, which specifies that there exists no measurement-error. Whereas, the sill-value is 17.397, which specifies that beyond this value there exists no correlation and this indicated that the variable is purely-random and the variogram flattens. The range-value is 90,431 for which the model initially flattens out.

The availability of phosphorus ranges between 6 kgha^{-1} –85kgha^{-1}. Applying kriging on the dataset with 685 observations for phosphorus-value, it predicts (left) and its associated kriging standard-error(right) is shown in Figure 25 for phosphorus-content. Then, kriging classifies the available-domain into different ranges like (22.7, 61.46), (61.46, 100.2), (100.2, 139), (139, 3.177.7), (177.7, 216.5), (216.5,255.2), (255.2,294) which is indicated by various colors on the map-area. The automatically created variogram as shown in Figure 24 and it is observed that the nugget-value is ‘0’ which specifies that no measurement-error exists. The sill-value is 1541, beyond which there exists no correlation and the variable is purely-random and the variogram flattens. The range-value is 87963 at which the model initially flattens-out.

The availability of potassium ranges between 88 kgha^{-1} – 324 kgha^{-1}. Applying kriging on the dataset with 685 observations for phosphorus-value, it predicts (left) and its associated kriging standard-error (right) is shown in Figure 26 for phosphorus-content. Then, kriging classifies the available-domain into different ranges like (8.457,53.52), (53.52,98.58), (98.58,143.6), (143.6,188.7), (188.7,233.8), (233.8,278.8), (278.8,323.9).The automatically created variogram as shown in Figure 26 and it is observed that the nugget-value is ‘0’ which specifies that no measurement-error exists. The sill-value is 26135, beyond which there exists no correlation and the variable is purely-random and the variogram flattens. The range-value is 83213 at which the model initially flattens-out.

After applying autokriging() on the defined soil-parameters (N, P, K) to the whole grid coordinates as shown in Figure 24, Figure 25 and Figure 26, it is observed that the positive range values suggest that (N, P, K) content is required in that zones. Whereas the negative range values suggest that (N, P, K) content is not required in that particular zone.

Figure 24. Kriging applied on Nitrogen

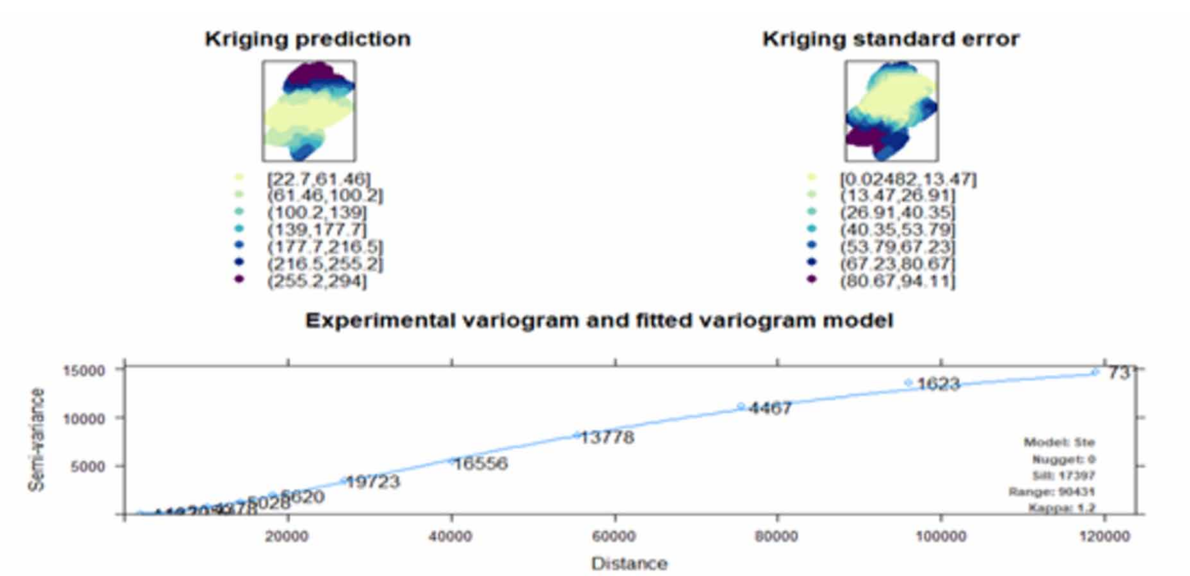


Figure 25. Kriging applied on phosphorus

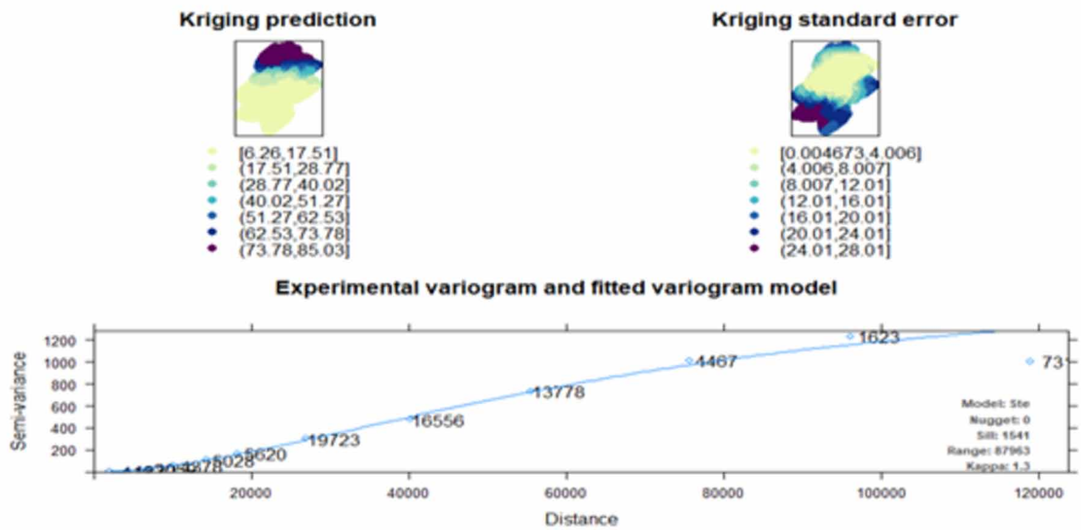
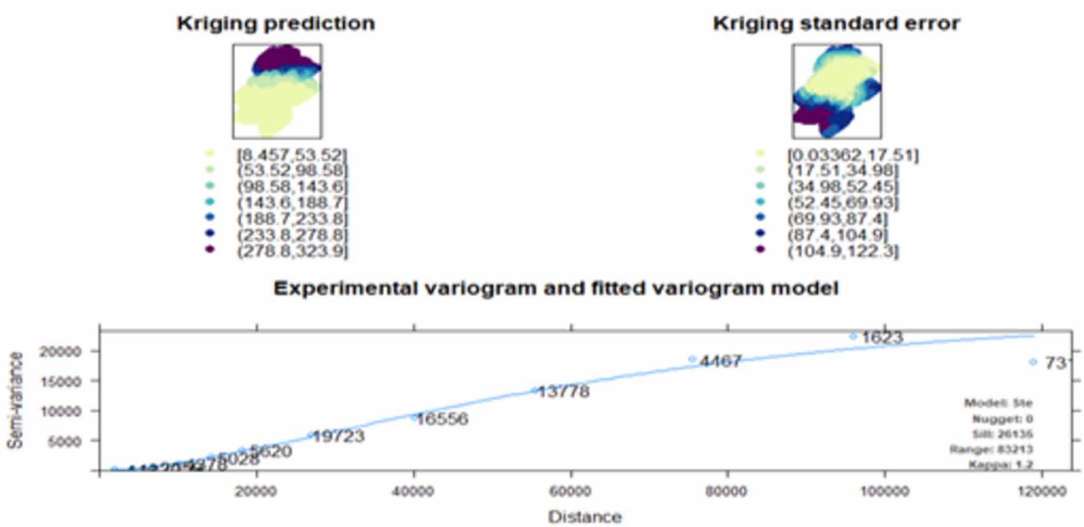


Figure 26. Kriging applied on potassium

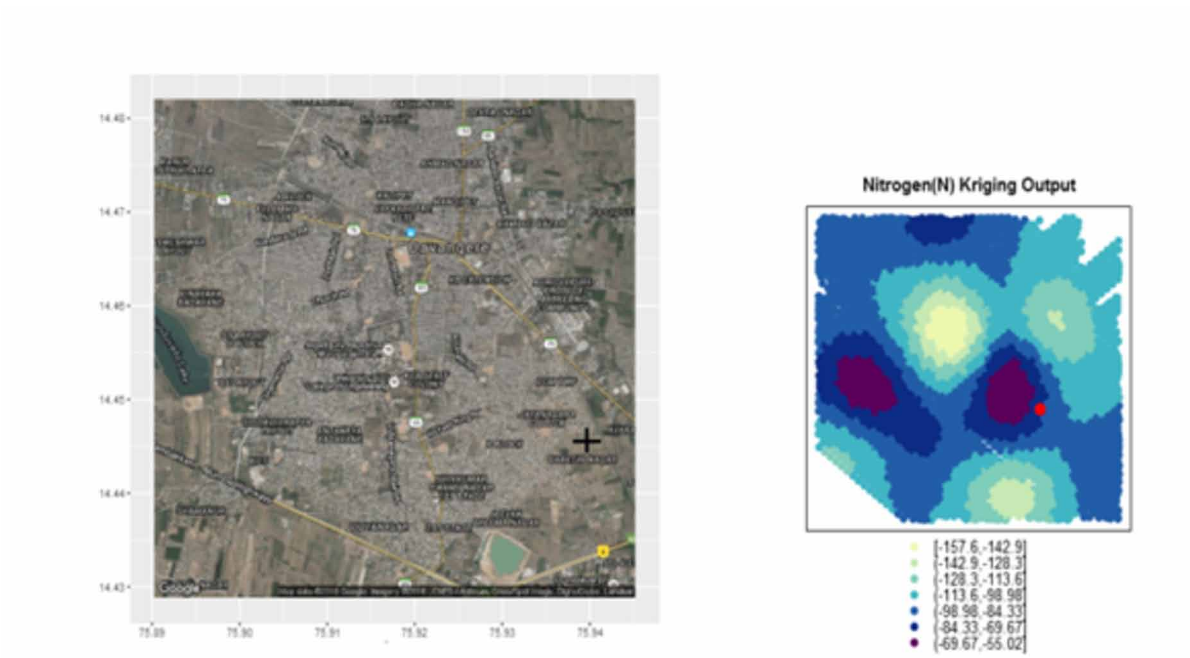


A user friendly and interactive Ajax based GUI is shown in Figure 27 is created for farmers using "R Shinng-App", that allows the farmer to select the location and soil-parameter with options as N, P, K, pH and all. Additional option for considering the temperature is also provided, where TR (Temperature Required and TNR (Temperature Not Required).

Figure 28 shows the fertilizer-map and the location-pointer over the Davangere-region on the Google-map. Here, the end-user can select any area-location point on the Davangere Google-map by clicking onthat particular point which is shown by the red-dot.Consider, for example user selects the area-location as "Shivaji Nagar, Davangere" to get then crop suggestion and the percentage of fertilizer required.

Figure 27. Front-End GUI for farmers

Figure 28. Kriging applied on the user selected area-location



DISCUSSION AND LIMITATIONS

For the given spatial-dataset, outliers can skew the conducted analysis and thus considering such outliers as true-observations and mistaking as clean-observations leads to masking/swamping effects. Hence, it is important to detect/remove such observations from the given dataset and reduce these effects. Iterative-Ration algorithm is applied which extends in multi-iterations and discovers one outlier in each iteration so that this outlier does not have negative impact on the next iterations. In-order to find the outliers, the graphical S-Outlier technique based on the spatial-data visualization is performed and it is observed that semi-variogram considers most of the N, K, P values that fall under the nugget-value as non-conforming values. But, these values are usually true-observations and not the outlying observations.

Further, Iterative-Ratio algorithm is applied for the detection/removal of outliers on the same given dataset, it is observed that the outliers found is more in number when compared to that of semi-variogram technique. From the obtained results, it is found that the semi-variogram mainly relies on the value of nugget-effect.

The soil-variables may gradually vary rather suddenly over-space thus de-limiting the spatial-boundaries of the given sub-field are explicitly is a difficult task. Here, from the analysis it is exploited that the spatial auto-correlation of such variables enhances the performance of k-means clustering algorithm for the sub-field zone-delineation. Thus kriging method is applied to treat the issues related to auto-correlation for the input-variable before performing the clustering-analysis. But, identifying the management-zones with its size is extremely related not only on the existing field-variations but also on the classification/clustering approaches.

The most widely used unconstrained fuzzy k-means algorithm when applied on the original-variables exhibits insufficiency for spatial auto-correlation. Hence, it is necessary for accounting the spatial-structure of the variables required for recognize the sub-field partitioning. The results obtained shows that combining k-means clustering with kriging presents a useful procedure for sub-field classification the input-variables. For the dataset considered in this work, spatial auto-correlation of the input-variables is stronger and mainly emphasize on the need to reflect auto-correlations for the study of spatial-variations. Moreover, the use of kriging and k-means let to identify the set of variables which accounted for global spatial-variances and of greatest spatial auto-correlation. By, applying MZD with the support of vario-gram allowed discovering the neighboring-points for the given data-points. The presented work, with the combination of kriging and k-means clustering convey that the obtained results are consistent with the other agronomic work discussed in the survey for delineating sub-field area with different yield-means.

Thus, the terrain/soil variables and significant auto-correlation presented high-loadings in the spatial-components. The mapping of kriging and k-means clustering discovered more contiguous-zones and finally resulted in highest-yield differences among the delineated zones and small within the class-variances.

The major limitations identified:

- Iterative-R algorithm as a binary classifier with sequential approach, when applied on large dataset becomes highly complex while evaluating its performance measure and thus it is computational intensive.
- For the given dataset, with various theoretical models presents different range-values that usually results in different management-zones and also affects the fertilizer variable rates for site-specific crop management.

- The soil-properties of the same given sample-field may exhibit different empirical semi-variogram models that have great impact on the data-analysis.

CONCLUSION AND FUTURE SCOPE

Precision Agriculture mainly focuses on accumulating huge amount of agricultural-data using various sensors to support crop-management decisions. The generated data usually consists of in-consistent data observation which deviates the crop-decisions assessments and also affects the yield-productivity. By, considering outliers as true-observations and misinterpreting as valid-observations as outliers is referred as swamping/masking effects. Thus, in this research work an attempt is made to reduce these effects in-line with the detection/removal of outliers from the given dataset by applying Iterative-R algorithm which discovers one-outlier in each iteration, so that the discovered outlier does not affect the consecutive iterations as negative. Further, outlier-free data points analyzed by Iterative-R algorithm are large when compared to that of the outlier data-points identified by semi-variogram. The results show that, for the given dataset with 853 observations with the actual range of N, K and P values only 214 observations are identified as complying-values (outlier free data).

In Precision Agriculture, the major issue lies in fertilization, which deals with availability of crop-minerals. Whereas, site-specific methods leads to imbalanced crop-management within farm-fields that affects the crop yield-productivity. Thus, the decision drawn by the farmers treats the entire field-area as uniform and forces them to use expensive resources like fertilizers/pesticides. For the existing heterogeneous field-area, it is necessary to identify the sub-fields and the percentage of fertilizers required. With the concern of improved yield-productivity the concept of MZD is adopted. The above outlier-free data is used by MZD to improve the decision on the crop-management and increase the yield-production. An interpolation method called kriging is used for MZD. The defined function conducts automatic-kriging on the given spatial-data and investigates the spatial-relationship among the data-records with the correlation-degree by comparing the nugget-to-sill ratio. The heterogeneity of the soil-parameters is evaluated on variance analysis. The fitted semi-variogram helps to define the size of variable-zones for the given field. The analysis carried out assists the farmers to locate the suitable farm-field area to grow the crops with minimal usage of fertilizers and to improve the crop-yield productivity. The work extends to reduce the environmental pollution risks with the over usage of fertilizers/pesticides.

Future Scope

Development of the future work focuses on strengthening the presented work by extending the investigation of efficiency of many other statistical techniques and graphical approaches

- Special attention can be given to examine the efficiency of multivariate spatial outlier algorithms to deal with the outliers in spatial-data specifically intended to support decision making process for precision-farming.
- For spatial-interpolation, further linear-regression model can be applied to improve the performance of distanced-based kriging.
- Extended attentiveness needs to be given for shape the spatial-structuring which largely discovers the interpolated-space and also prediction-errors.

- More efficient techniques are required to form the semi-variogram when availability of data is limited.
- The Prediction Accuracy and Efficiency of outlier detection methods can be measured by applying confusion matrix and ROC curve.

ACKNOWLEDGMENT

The research work was supported and funded by Vision Group of Science and Technology (VGST-K-FIST Level-I), Bengaluru, INDIA. [Grant number: 338] with amount of 40 Lakhs to meet the technological and Scientific needs. An extended thanks for the technical staff of TarabaluKrishi Kendra for supporting and providing the necessary dataset of the study-area (entire Davangere District). Finally, thanks to the Institution and the management for providing the necessary technical support to carry-out the research work successfully.

REFERENCES

Biswas & Si. (2013). Model Averaging for Semi-variogram Model Parameters. Advances in Agrophysical Research, 81-94.

Cai, He, & Mana. (2013). Spatial outlier detection based on iterative-self-organizing learning model. *Neurocomputing*, *117*, 61–172.

Cárdenas-Montes, M. (2014). Depth-Based Outlier Detection Algorithm. M. *Springer International Publishing Switzerland.*, *84840*, 122–132.

Córdoba, M., Bruno, C., Costa, J., & Balzarini, M. (2013). Subfield management class delineation using cluster analysis from spatial principal components of soil variables. *Computers and Electronics in Agriculture*, *97*, 6–14. doi:10.1016/j.compag.2013.05.009

Dik, Jebari, & Bouroumi. (2014). Similarity-based approach for outlier detection. *International Journal of Computer Science, 11*, 41-45.

Dimble, N. M., & Tidke, B. (2015). A Framework for outlier Detection in Geographic Spatial data. *International Journal of Foundations of Computer Science*, *2*, 59–67.

Guastaferro, F., Castrignanò, A., Benedetto, D., Sollitto, D., Troccoli, A., & Cafarelli, B. (2010). A comparison of different algorithms for the delineation of management zones. *Precision Agriculture*, *11*(6), 600–620. doi:10.100711119-010-9183-4

Guvendik, Genc, Tamer, & Nil. (2012). Improving the performance of Kriging based interpolation application with parallel processors. In *Signal Processing and Communications Applications Conference (SIU).* IEEE.

Horvath & Zelenika. (2011). *Identifcation of palaeoenvironments using clustering methods and indicator kriging, Case study from Late Miocene sandstones, the Sava depression*. Academic Press.

Kruize, Wolfert, Scholten, Verdouw, Kassahun, & Beulens. (2016). *A Reference architecture for Farm Software Ecosystems*. Academic Press.

Lokers, Knapen, Janssen, van Randen, & Jansen. (2016). *Analysis of Big Data technologies for use in agro-environmental science.* Academic Press.

Lu, C.-T., Chen, D., & Kou, Y. (2003). Algorithms for Spatial Outlier Detection. *Third IEEE International Conference on Data Mining (ICDM'03)*. 10.1109/ICDM.2003.1250986

Majumdar, Sneha, & Shilpa. (2017). Analysis of agriculture data using data mining techniques: application of big data. *Journal of Big Data*, 4-20.

Peralta, Costa, Balzarini, Franco, Córdoba, & Bullock. (2015). Delineation of management zones to improve nitrogen management of wheat. *Computers and Electronics in Agriculture, 110*, 103-113.

Rub & Brenning. (2010). *Data-mining in precision agriculture: Management of spatial information.* Academic Press.

Rub & Kruse. (2011). Exploratory hierarchical clustering for management zone delineation in precision agriculture. In *Advances in Data Mining. Applications and Theoretical Aspects*. Springer Berlin Heidelberg.

Sabarish, B. A., & Karthi, R. (2018). Spatial Outlier Detection Algorithm for Trajectory-Data. *International Journal of Pure and Applied Mathematics*, *118*, 325–331.

Serfling & Wang. (2012). *General foundations for studying masking and swamping robustness of outlier identifiers*. Academic Press.

Shekhar, Vatsavai, & Celik. (2012). Spatiotemporal data mining in the era of big spatial data: algorithms and applications. *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data*, 1-10.

Su. (2011). Statistical Geo-computing: Spatial Outlier Detection in Precision Agriculture. Academic Press.

Taechatanasat & Armstrong. (2014). Decision Support System Data for Farmer Decision Making. *Proceedings of Asian Federation for Information Technology in Agriculture*, 472-486.

Wang, Xia, & Wang. (2016). A *Spectral Clustering Basedoutlier detection technique*. Springer International Publishing.

Wolfert, Ge, Verdouw, & Bogaardt. (2017). Big Data in Smart Farming–A review. *Agricultural System*, (153), 69-80.

Zhang, Jiang, Qiu, Qiu, Wang, & Zhu. (2017). An improved method of delineating rectangular management zones using a semivariogram-based technique. Computers and Electronics in Agriculture, 121, 74-83.

Compilation of References

Abdullah, A., Brobst, S., Pervaiz, I., Umer, M., & Nisar, A. (2004). Learning dynamics of pesticide abuse through data mining. *Proc. of Australian Workshop on Data Mining and Web Intelligence*.

Abello, J., Pardalos, P. M., & Resende, M. (2002). *Handbook of massive data sets*. New York: Kluwer. doi:10.1007/978-1-4615-0005-6

Adamchuk, V. I., Hummel, J. W., Morgan, M. T., & Upadhyaya, S. K. (2004). On-the-go soil sensors for precision agriculture. *Computers and Electronics in Agriculture*, *44*(1), 71–91. doi:10.1016/j.compag.2004.03.002

Aggarwal, M., Kaushik, A., Sengar, A., Gangwar, A., Singh, A. K., & Raj, V. (2014). Agro App: An application for healthy living. *2014 International Conference on Information Systems and Computer Networks (ISCON)*, 30-32 10.1109/ICISCON.2014.6965213

Agus, F., Ihsan, M., Khairina, D. M., & Candra, K. P. (2019). *Expert System for Rice Plant Disease Diagnosis*. Retrieved from https://www.researchgate.net/publication/331260352

Ahmed, S. R. (2004). Applications of data mining in retail business. *International Conference on Information Technology: Coding Computing, 2*, 455-459. 10.1109/ITCC.2004.1286695

Aibinu, A. M., Salami, M. J. E., Shafie, A. A., Hazali, N., & Termidzi, N. (2011, January). Automatic fruits identification system using hybrid technique. In *Electronic Design, Test and Application (DELTA), 2011 Sixth IEEE International Symposium on* (pp. 217-221). IEEE. 10.1109/DELTA.2011.47

Al-Hiary, H., Bani-Ahmad, S., Reyalat, M., & Braik, M., & Al-Rahamneh, Z. (2011). Fast and accurate detection and classification of plant diseases. *Machine Learning*, *14*(5).

Ali & Kapoor. (2019). Farmers' perception on risks in fruits and vegetables production: an empirical study of Uttar Pradesh. *Agricultural Economics Research Review*, 21.

Ali, Sarah. (2014). Image Enhancement -Spatial vs. Frequency Domain Filters. . doi:10.13140/2.1.1256.7042

Ammad-udin, M., Mansour, A., Le Jeune, D., Aggoune, E. H. M., & Ayaz, M. (2016). *"UAV routing protocol for crop health management," 2016 24th European Signal Processing Conference* (pp. 1818–1822). Budapest: EUSIPCO; doi:10.1109/EUSIPCO.2016.7760562.

Amrutha, A., Lekha, R., & Sreedevi, A. (2016). *Automatic Soil Nutrient Detection and Fertilizer Dispensary System*. IEEE. doi:10.1109/RCTFC.2016.7893418

Andujar, M., Aroba, J., de Torre, M. L., & Grande, J. A. (2006). Contrast of evolution models for agricultural contaminants in ground waters by means of fuzzy logic and data mining. *Environmental Geology*, *49*(3), 458–466. doi:10.100700254-005-0103-2

Angadi & Hatture. (2011). A novel spectral graph theoretic approach to user identification using hand geometry. *International Journal of Machine Intelligence, 3*(4), 282-288.

Angadi & Hatture. (2019). Face Recognition Through Symbolic Modeling of Face Graphs and Texture. *International Journal of Pattern Recognition and Artificial Intelligence*. doi:10.1142/S0218001419560081

Angadi, S. A., & Hatture, S. M. (2018). Hand Geometry Based User Identification Using Minimal Edge Connected Hand Image Graph. *IET Computer Vision*, *12*(5), 744–752. doi:10.1049/iet-cvi.2017.0053

An, N., Palmer, C. M., Baker, R. L., Markelz, R. C., Ta, J., Covington, M. F., & Weinig, C. (2016). Plant high-throughput phenotyping using photogrammetry and imaging techniques to measure leaf length and rosette area. *Computers and Electronics in Agriculture*, *127*, 376–394. doi:10.1016/j.compag.2016.04.002

Aradhya. (2007). Language Independent Skew Estimation Technique Based on Gaussian Mixture Models: A Case Study on South Indian Scripts. *Lecture Notes in Computer Science*.

Arbelaez, P., Maire, M., Fowlkes, C., & Malik, J. (2011). Contour detection and hierarchical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*(5), 898–916. doi:10.1109/TPAMI.2010.161 PMID:20733228

Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J. M., & Perona, I. (2013). An extensive comparative of cluster validity indices. *Pattern Recognition*, *46*(1), 243–256. doi:10.1016/j.patcog.2012.07.021

Arefi, A., Motlagh, A. M., Mollazade, K., & Teimourlou, R. F. (2011). Recognition and localization of ripen tomato based on machine vision. *Australian Journal of Crop Science*, *5*(10), 1144.

Arthur, D., & Vassilvitskii, S. (2006, June). How slow is the k-means method? In *Proceedings of the twenty-second annual symposium on Computational geometry* (pp. 144-153). ACM. 10.1145/1137856.1137880

Asghari, P., Rahmani, A. M., & Javadi, H. H. S. (2019). Internet of Things applications: A systematic review. *Computer Networks*, *148*, 241–261. doi:10.1016/j.comnet.2018.12.008

Ashour, A. S., Samanta, S., Dey, N., Kausar, N., Abdessalemkaraa, W. B., & Hassanien, A. E. (2015). Computed Tomography Image Enhancement Using Cuckoo Search: A Log Transform Based Approach. *Journal of Signal and Information Procedureing*, *6*(03), 244–257. doi:10.4236/jsip.2015.63023

Ashourloo, D., Matkan, A. A., Huete, A., Aghighi, H., & Mobasheri, M. R. (2016). Developing an index for detection and identification of disease stages. *IEEE Geoscience and Remote Sensor Letters*, *13*(6), 851–855.

Astonkar, S. R., & Shandilya, V. K. (2018). Detection and analysis of plant diseases using image processing technique. International Research Journal of Engineering and Technology (IRJET), 05(04), 4. Retrieved from https://www.irjet.net/archives/V5/i4/IRJET-V5I4717.pdf

Athani, S., Tejeshwar, C. H., Patil, M. M., Patil, P., & Kulkarni, R. (2017). Soil moisture monitoring using IoT enabled Arduino sensors with neural networks for improving soil management for farmers and predict seasonal rainfall for planning future harvest in North Karnataka — India. In *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)* (pp. 43–48). Palladam, Tamilnadu, India: IEEE. 10.1109/I-SMAC.2017.8058385

Avendano, J., Ramos, P. J., & Prieto, F. A. (2017). A system for classifying vegetative structures on coffee branches based on videos recorded in the field by a mobile device. Expert Systems with Applications, 88, 178e192.

Babinet, G. (2015). *The New World economy, s.l.: Report addressed to Ms Segolene Royal, Minister of Environment, Sustainable Development and Energy.* Working group led by Corinne Lepage.

Babu, Ramana Murthy, & Narayana. (2010). A web based tomato crop expert information system based on artificial intelligence and machine learning algorithms. *Int. J. of Comp. Sci., and Information Technologies*, *1*(1), 6-15.

Bah, M.D., Hafiane, A., & Canals, R. (2018). *Deep Learning with unsupervised data labeling for weeds detection on UAV images.* CoRR, abs/1805.12395

Bajwa, S., Bajcsy, P., Groves, P., & Tian, L. F. (2004). Hyperspectral image data mining for band selection in agricultural applications. *Transactions of the ASAE. American Society of Agricultural Engineers*, *47*(3), 895–907. doi:10.13031/2013.16087

Bakhshinategh, B., Zaiane, O. R., ElAtia, S., & Ipperciel, D. (2018). Educational data mining applications and tasks: A survey of the last 10 years. *Education and Information Technologies*, *2*(1), 537–553. doi:10.100710639-017-9616-z

Barbedo, J. (2014). Using Digital Image Processing for Counting Whiteflies on Soybean Leaves. *Journal of Asia-Pacific Entomology*, *17*(4), 685–694. doi:10.1016/j.aspen.2014.06.014

Bargoti, S., & Underwood, P. James. (2016). Image Segmentation for Fruit Detection and Yield Estimation in Apple Orchards. *Journal of Field Robotics*. doi:10.1002/rob.21699

Barkunan, S. R., Bhanumathi, V., & Sethuram, J. Smart sensor for automatic drip irrigation system for paddy cultivation, Computers & Electrical Engineering, Volume 73, 2019, Pages 180-193, ISSN 0045-7906, doi:10.1016/j.compeleceng.2018.11.013

Basak, J., Sudharshan, A., Trivedi, D., & Santhanam, M. S. (2004). Weather Data Mining Using water/nitrogen budgets and crop yield. *Agricultural Systems*, *46*(3), 335–359.

Baskar, S. S., Arockiam, L., & Charles, S. (2013). *Applying Data Mining Techniques on Soil Fertility Prediction*. *IJCAT*. doi:10.7753/IJCATR0206.1005

Bastiaanssen, W., Molden, D., & Makin, I. (2000). Remote sensing for irrigated agriculture: Examples from research and possible applications. *Agricultural Water Management*, *46*(2), 137–155. doi:10.1016/S0378-3774(00)00080-9

Beniwal, S., & Arora, J. (2012). Classification and feature selection techniques in data mining. *International Journal of Engineering Research & Technology*, *1*(6).

Berry, M. J. A., & Linoff, G. S. (2004). *Data mining technique for marketing, sales, and customer relationship management* (2nd ed.). Wiley.

Bhagawati, R., Bhagawati, K., Singh, A. K. K., Nongthombam, R., Sarmah, R., & Bhagawati, G. (2015). Artificial Neural Network Assisted Weather Based Plant Disease Forecasting System. *International Journal on Recent and Innovation Trends in Computing and Communication*, *3*(6), 4168–4173.

Bhange, M., & Ali, H. (2015). Smart Farming: Pomegranate Disease Detection Using Image Processing. *Procedia Computer Science*, *58*, 280–288. doi:10.1016/j.procs.2015.08.022

Bhargavi, P, & Jyothi, S. (2009). Applying Naive Bayes data mining technique for classification of agricultural land soils. *International Journal of Computer Science and Network Security*, *9*(8), 117-122.

Bhasin, M. L. (2006). Data mining: A competitive tool in the banking and retail industries. ML Bhasin. *Chartered Accountant-New Delhi*, *55*(4), 588.

Bhuyar, V. (2014). *Comparative Analysis of Classification Techniques on Soil Data to Predict Fertility Rate for Aurangabad District*. *IJETTCS*.

Biradar, H. B., & Shabadi, L. (2017). Review on IOT based multidisciplinary models for smart farming. In *2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)* (pp. 1923–1926). Bangalore: IEEE. 10.1109/RTEICT.2017.8256932

Bishop, C. (2006). *Pattern Recognition and Machine Learning*. New York: Springler-Verlag.

Biswas & Si. (2013). Model Averaging for Semi-variogram Model Parameters. Advances in Agrophysical Research, 81-94.

Bonanomi, G., Antignani, V., Capodilupo, M., & Scala, F. (2010). Identifying the characteristics of organic soil amendments that suppress soilborne plant diseases. *Soil Biology & Biochemistry*, *42*(2), 136–144. doi:10.1016/j.soilbio.2009.10.012

Bongiovanni, R., & Lowenberg-DeBoer, J. (2004). Precision agriculture and sustainability. *Precision Agriculture*, *5*(4), 359–387. doi:10.1023/B:PRAG.0000040806.39604.aa

Braimoh, Vlek, & Stein. (2004). Land Evaluation for Maize Based on Fuzzy Set and Interpolation. *Environmental Management*, *33*(2), 236–238. PMID:15285400

Braimoh, A. K., & Vlek Paul, L. (2004). Land Evaluation for Maize Based on Fuzzy Set and Interpolation. *Environmental Management*, *33*(2), 226–238. doi:10.100700267-003-0171-6 PMID:15285400

Brajovic, Vujovic, & Dukanovic. (2015). An Overview of Smart Irrigation Software. *4th Mediterranean Conference on Embedded Computing, MECO*.

Brazilian Agricultural Research Corporation. Visão 2014–2034. (2014). Retrieved May 17, 2019, from Link: https://www.embrapa.br/documents/1024963/1658076/Documento+Visão+-+versão+completa/7bf520f2-7329-42c0-8bf0-15b3353c3fdb

Brettel, Friederichsen, Keller, & Rosenberg. (2014). How virtualization, decentralization and network building change the manufacturing landscape: an industry 4.0 perspective *IJMEA*, *8*(1), 37-44.

Cai, He, & Mana. (2013). Spatial outlier detection based on iterative-self-organizing learning model. *Neurocomputing*, *117*, 61–172.

Camilli, A., Cugnasca, C. E., Saraiva, A. M., Hirakawa, A. R., & Corrêa, P. L. P. (2007). From wireless sensors to field mapping: Anatomy of an application for precision agriculture. *Computers and Electronics in Agriculture*, *58*(1), 25–36. doi:10.1016/j.compag.2007.01.019

Camps-Valls, G., Gomez-Chova, L., Calpe-Maravilla, J., Soria-Olivas, E., Martin-Guerrero, J. D., & Moreno, J. (2003). Support vector machines for crop classification using hyperspectral data. *Lecture Notes in Computer Science*, *2652*, 134–141. doi:10.1007/978-3-540-44871-6_16

Cárdenas-Montes, M. (2014). Depth-Based Outlier Detection Algorithm. M. *Springer International Publishing Switzerland.*, *84840*, 122–132.

Car, N., Christen, E., Hornbuckle, J., & Moore, G. (2012). Using a mobile phone Short Messaging Service (SMS) for irrigation scheduling in Australia – Farmers' participation and utility evaluation. *Computers and Electronics in Agriculture*, *84*, 132–143. doi:10.1016/j.compag.2012.03.003

Cerutti, G., Tougne, L., Vacavant, A., & Coquin, D. (2011, September).A parametric active polygon for leaf segmentation and shape estimation. In *Proceedings of International Symposium on Visual Computing* (pp. 202-213). Springer. 10.1007/978-3-642-24028-7_19

Challa, M. (2013). *Determining Factors and Impacts of Modern Agricultural Technology Adoption in West Wollega, Munich*. GRIN Publishing GmbH. Retrieved May 15, 2019, from://www.grin.com/en/e-book/280336/determiningfactors-and-impacts-of-modern-agricultural-technology-adoption

Chandola, V., Banerjee, A., & Vipin, K. (2009). Anomaly detection: A survey. *ACM Computing Surveys*, *41*(3), 1–58. doi:10.1145/1541880.1541882

Chandra. (2011). Role of E-Agriculture in Rural Development in Indian Context. *International Conference on Emerging Trends in Networks and Computer Communications (ETNCC)*.

Chaochong, J. M. (2008). *Forecasting Agricultural Production via Generalized Regression Neural Network*. IEEE.

Chen, S.W., Shivakumar, S.S., Dcunha, S., Das, J., Okon, E., Qu, C., Taylor, C.J. & Kumar. (2017). Counting apples and oranges with deep learning: A data-driven approach. *IEEE Robotics and Automation Letters*, *2*(2), 781-788.

Chéné, Y., Rousseau, D., Lucidarme, P., Bertheloot, J., Caffier, V., Morel, P., & Chapeau-Blondeau, F. (2012). On the use of depth camera for 3D phenotyping of entire plants. *Computers and Electronics in Agriculture*, *82*, 122–127. doi:10.1016/j.compag.2011.12.007

Chi, M., Plaza, A., Benediktsson, J. A., Sun, Z., Shen, J., & Zhu, Y. (2016). Big data for remote sensing: Challenges and opportunities. *Proceedings of the IEEE*, *104*(11), 2207–2219. doi:10.1109/JPROC.2016.2598228

Choudhary, R., & Parmar, G. (2016, November).A robust image watermarking technique using 2-level discrete wavelet transform (DWT). In *Proceedings of 2016 2nd International Conference on Communication Control and Intelligent Systems (CCIS)*, (pp. 120-124). IEEE. 10.1109/CCIntelS.2016.7878213

Chunhua Zhu, Jiaojiao Wang, Hao Liu, and Huan Mi, "Insect Identification and Counting in Stored Grain: Image Processing Approach and Application Embedded in Smartphones," Mobile Information Systems, vol. 2018, Article ID 5491706, 5 pages, 2018. . doi:10.1155/2018/5491706

Clement, J., Novas, N., Gázquez, J. A., & Manzano-Agugliaro, F. (2012). High Speed Intelligent Classifier Of Tomatoes By Colour, Size And Weight. *Spanish Journal of Agricultural Research*, *10*(2), 314–325. doi:10.5424jar/2012102-368-11

Collins, R. J. (2009). *Value Chain Management And Postharvest Handling: Partners In Competitiveness*. Academic Press.

Córdoba, M., Bruno, C., Costa, J., & Balzarini, M. (2013). Subfield management class delineation using cluster analysis from spatial principal components of soil variables. *Computers and Electronics in Agriculture*, *97*, 6–14. doi:10.1016/j.compag.2013.05.009

Corwin, D. L., & Lesch, S. M. (2005). Apparent soil electrical conductivity measurements in agriculture. *Computers and Electronics in Agriculture*.

Cox, S. (2002). Information technology: The global key to precision agriculture and sustainability. *Computers and Electronics in Agriculture*, *36*(2-3), 93–111. doi:10.1016/S0168-1699(02)00095-9

Cunningham & Holmes. (1999). *Developing innovative applications in agriculture using data mining*. Academic Press.

Dadwal, M., & Banga, V. (2012, October). Estimate Ripeness Level of fruits Using RGB Color Space and Fuzzy Logic Technique. *International Journal of Engineering and Advanced Technology*, *2*(1), 225–229.

Dang, H., Song, J., & Guo, Q. (2010, August). A Fruit Size Detecting And Grading System Based On Image Processing. In *Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2010 2nd International Conference on* (Vol. 2, pp. 83-86). IEEE. . 12010.1109/IHMSC.2010.120

Das, S., & Nayak, L. (2013). Expert System for Agriculture Extension. *Indian Journal of Automation and Artificial Intelligence, 1*(2). Retrieved from http://ijaai.informaticspublishing.com/index.php/ijaai/article/view/32577

Dass. (n.d.). *Data Mining in Banking and Finance: A Note for Bankers*. Indian Institute of Management Ahmadabad.

De Vylder, J., Vandenbussche, F. J., Hu, Y., Philips, W., & Van Der Straeten, D. (2012). Rosette tracker: an open source image analysis tool for automatic quantification of genotype effects. *Plant Physiology*, 112.

De Vylder, J., Ochoa, D., Philips, W., Chaerle, L., & Van Der Straeten, D. (2011, October).Leaf segmentation and tracking using probabilistic parametric active contours. In *Proceedings of International Conference on Computer Vision/ Computer Graphics Collaboration Techniques and Applications* (pp. 75-85). Springer. 10.1007/978-3-642-24136-9_7

Deepika, G., & Rajapirian, P. (2016). Wireless sensor network in precision agriculture: A survey. In *2016 International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS)* (pp. 1–4). Pudukkottai, India: IEEE. 10.1109/ICETETS.2016.7603070

Deepthi, M. B., & Sreekantha, D. K. (2017). Application of Expert Systems for Agricultural Crop Disease Diagnoses - A Review. *International Conference on Inventive Communication and Computational Technologies (ICICCT 2017)*. Retrieved from https://ieeexplore.ieee.org/iel7/7966452/7975154/07975192.pdf

DEFRA. (2013). Farm Practices Survey Autumn 2012 – England. Department for Environment, Food and Rural Affairs (DEFRA), 41.

Dellen, B., Scharr, H., & Torras, C. (2015). Growth signatures of rosette plants from time-lapse video. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *12*(6), 1470–1478. doi:10.1109/TCBB.2015.2404810 PMID:26684463

Department of Horticulture. (2014). *Post Harvest Technology of Horticultural Crops. HISAR*. Haryana, India: CCS Haryana Agricultural University.

Desai & Desai. (2004). The Role of Data mining in Banking Sector. *IBA Bulletin*.

Dey, K., Sharma, M., & Meshram, M. R. Image processing based leaf rot disease, detection of betel vine (Piper BetleL.), in: Proceedings of the International Conference on Computational Modeling and Security (CMS), 2016, pp. 748–754. doi:10.1016/j.procs.2016.05.262

Dhole, S. A., & Shaikh, R. P. (2016). Review of leaf unhealthy region detection using image processing techniques. *Bulletin of Electrical Engineering and Informatics*, *5*(4), 3. doi:10.11591/eei.v5i4.498

Dik, Jebari, & Bouroumi. (2014). Similarity-based approach for outlier detection. *International Journal of Computer Science, 11*, 41-45.

Dimble, N. M., & Tidke, B. (2015). A Framework for outlier Detection in Geographic Spatial data. *International Journal of Foundations of Computer Science*, *2*, 59–67.

Dimitriadis, S., & Goumopoulos, C. (2008). Applying Machine Learning to Extract New Knowledge in Precision Agriculture Applications. In *2008 Panhellenic Conference on Informatics* (pp. 100–104). Samos, Greece: IEEE. 10.1109/PCI.2008.30

Ding, L., Meng, J., & Yang, Z. (2010). An Early Warning System of Pork Price in China Based on Decision Tree. *IEEE Conference Proceedings, International Conference on E-Product E-Service and EEntertainment*, 1–6. 10.1109/ICEEE.2010.5660808

Dorj, U.-O., Lee, M., & Yun, S. (2017). An yield estimation in citrus orchards via fruit detection and counting using image processing. *Computers and Electronics in Agriculture*, *140*, 103–112. doi:10.1016/j.compag.2017.05.019

Dr.K.Thangadurai, K.Padmavathi, "Computer Vision image Enhancement For Plant Leaves Disease Detection", 2014 World Congress on Computing and Communication Technologies.

Dubey, Kamath, & Kanakia. (2016). Learning Data Mining Techniques. *International Journal of Computer Applications*.

Dwivedy. (2011). Challenges faced by agriculture sector on developing countries with special reference India. *International Journal of Rural Studies, 18*(2).

Eapen. (2004). *Application of Data mining in Medical Applications*. UWSpace. Retrieved from http://hdl.handle.net/10012/772

Effendi, Z., Ramli, R., Ghani, J. A., & Rahman, M. N. A. (2009, November). Pattern Recognition System Of Jatropha-Curcas Fruits Using Back Propagation. In *Signal and Image Processing Applications (ICSIPA), 2009 IEEE International Conference on* (pp. 58-62). IEEE. doi: 10.1109/ICSIPA.2009.5478719

Elemo, T. K., & Kumar, M. (2018). Diagnosis System for Disease in Wheat plant using Intelligent Expert System: Design and Implementation. *International Journal of Pure and Applied Mathematics, 118*(24). Retrieved from http://www.acadpubl.eu/hub/

Elizondo, D. A., McClendon, R. W., & Hoogenboom, G. (1994). Neural network models for predicting flowering and physiological maturity of soybean. *Transactions of the ASAE. American Society of Agricultural Engineers*.

Emre, I. E., Erol, N., Ayhan, Y. İ., Özkan, Y., & Erol, Ç. (2019). The analysis of the effects of acute rheumatic fever in childhood on cardiac disease with data mining. *International Journal of Medical Informatics*, *123*, 68–75. doi:10.1016/j.ijmedinf.2018.12.009 PMID:30654905

Everitt, B., Landau, S., & Leese, M. (2001). *Cluster Analysis*. London: Arnold.

Fayyad, Piatetsky-Shapiro, & Smyth. (1996). *Advances in knowledge discovery and data mining*. Academic Press.

Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture*, *145*, 311–318. doi:10.1016/j.compag.2018.01.009

Font Calafell, D. (2014). An Image Processing Method For In-Line Nectarine Variety Verification Based On The Comparison Of Skin Feature Histogram Vectors. *Computers and Electronics in Agriculture*, *102*, 112–119. doi:10.1016/j.compag.2014.01.013

Font, D., Tresanchez, M., Mart'ınez, D., Moreno, J., Clotet, E., & Palac'ın, J. (2015). Vineyard yield estimation based on the analysis of high resolution images obtained with artificial illumination at night. *Sensors (Basel)*, *15*(4), 8284–8301. doi:10.3390/s150408284 PubMed

Food & Agricultural Organisation. (2009). *How to Feed the World in 2050*. Retrieved from http://www.fao.org/fileadmin/templates/wsfs/docs/expert_paper/How_to_Feed_the_World_in_2050.pdf

Forkuor, G., Landmann, T., Conrad, C., & Dech, S. W. (2012). Agricultural land use mapping in the sudanian savanna of West Africa: Current status and future possibilities. *2012 IEEE International Geoscience and Remote Sensing Symposium*, 6293-6296. 10.1109/IGARSS.2012.6352698

Fountas, S., Pedersen, S.M., & Blackmore, S. (2005). ICT in Precision Agriculture – diffusion of technology. ICT in agriculture: Perspectives of Technological Innovation, 15.

Fountas, S., Carli, G., Sørensen, C. G., Tsiropoulos, Z., Cavalaris, C., & Vatsanidoud, A. (2015). Farm management information systems: Current situation and future perspectives. *Computers and Electronics in Agriculture*, *115*, 40–50. doi:10.1016/j.compag.2015.05.011

Fu, Longsheng & Liu, Zhihao & Majeed, Yaqoob & Cui, Yongjie. (2018). Kiwifruit yield estimation using image processing by an Android mobile phone. 51. 185-190. . doi:10.1016/j.ifacol.2018.08.137

Gajanan, D.E., Shankar, G., & Keshav, G.V. (2018). Detection of Leaf Disease Using Feature Extraction for Android Based System.

Gandge, Y. & Sandhya. (2017). A study on various data mining techniques for crop yield prediction. *International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT)*, 420-423. 10.1109/ICEECCOT.2017.8284541

Gandhamal, A., Talbar, S., Gajre, S., Hani, M., Fadzil, A., & Kumar, D. (2017). Local Gray Level S-Curve Transformation – A Generalized Contrast Enhancement Technique for Medical Images. *Computers in Biology and Medicine*, *83*, 120–133. doi:10.1016/j.compbiomed.2017.03.001 PubMed

Gandhi, R., Nimbalkar, S., Yelamanchili, N., & Ponkshe, S. (2018). Plant disease detection using CNNs and GANs as an augmentative approach. In *2018 IEEE International Conference on Innovative Research and Development (ICIRD)* (pp. 1–5). Bangkok: IEEE. 10.1109/ICIRD.2018.8376321

Gaurav, Viveka, Rajesh, & Shaila. (2018, February). Automated Number Plate Verification System based on Video Analytics. *International Research Journal of Engineering and Technology*.

Gayatri, M., Jayasakthi, J., & Mala, G.S. (2015). Providing Smart Agricultural solutions to farmers for better yielding using IoT. *2015 IEEE Technological Innovation in ICT for Agriculture and Rural Development (TIAR)*, 40-43.

Gebbers, R., & Adamchuk, V. (2010). Precision agriculture and food security. *Science*, *327*(5967), 828–831. doi:10.1126cience.1183899 PMID:20150492

Gers, F., Schmidhuber, J., & Cummins, F. (1999). Learning to forget: continual prediction with LSTM. In *1999 Ninth International Conference on Artificial Neural Networks ICANN 99*. Edinburgh, UK: IET, . 10.1049/cp:19991218

Gerven, M., & Bohte, S. (Ed.). (2017). Artificial Neural Networks as Models of Neural Information Processing. Frontiers' Research Topics.

Ghadiyali, T., Lad, D. K., & Patel, B. (2011). Agriculture Intelligence: An Emerging Technology for Farmer Community. *2011 Second International Conference on Emerging Applications of Information Technology*, 313-316. 10.1109/EAIT.2011.36

Gholap, J. (2012). Performance tuning of j48 algorithm for prediction of soil fertility. *Asian Journal of Computer Science and Information Technology*, *2*(8), 251–252.

Gill, J., Sandhu, D. S., & Singh, D. (2014). A Review of Automatic Fruit Classification using Soft Computing Techniques. In *International Conference on Computer, Systems and Electronics Engineering (ICSCEE'2014),* (pp. 91-98). Johannesburg, South Africa: Grand View Research. Retrieved May 17, 2019, from https://www.grandviewresearch.com/press-release/global-agriculture-drones-market

Giuffrida, M. V., Minervini, M., & Tsaftaris, S. A. (2016). *Learning to count leaves in rosette plants*. Academic Press.

Goh, A. T. C. (1995). Back-propagation neural networks for modeling complex systems. *Artificial Intelligence in Engineering*, *9*(3), 143–151. doi:10.1016/0954-1810(94)00011-S

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

Grand-Brochier, M., Vacavant, A., Cerutti, G., Kurtz, C., Weber, J., & Tougne, L. (2015). Tree leaves extraction in natural images: Comparative study of preprocessing tools and segmentation methods. *IEEE Transactions on Image Processing*, *24*(5), 1549–1560. doi:10.1109/TIP.2015.2400214 PMID:25667351

Grossman, R., Kamath, C., Kegelmeyer, P., Kumar, V., & Namburu, R. (2001). *Data Mining for Scientific and Engineering Applications*. doi:10.1007/978-1-4615-1733-7

Guastaferro, F., Castrignanò, A., Benedetto, D., Sollitto, D., Troccoli, A., & Cafarelli, B. (2010). A comparison of different algorithms for the delineation of management zones. *Precision Agriculture*, *11*(6), 600–620. doi:10.100711119-010-9183-4

Guptaa, S., & Trivedi, G. (2016). e-krishakMitra. *International Conference on Accessibility to Digital World (ICADW)*.

Gutiérrez, J., Villa-Medina, J. F., Nieto-Garibay, A., & Porta-Gándara, M. Á. (2014, January). Automated Irrigation System Using a Wireless Sensor Network and GPRS Module. *IEEE Transactions on Instrumentation and Measurement*, *63*(1), 166–176. doi:10.1109/TIM.2013.2276487

Guvendik, Genc, Tamer, & Nil. (2012). Improving the performance of Kriging based interpolation application with parallel processors. In *Signal Processing and Communications Applications Conference (SIU)*. IEEE.

Haff, R. P., & Toyofuku, N. (2008). X-ray detection of defects and contaminants in the food industry. *Sensing and Instrumentation for Food Quality and Safety*, *2*(4), 262–273. doi:10.1007/s11694-008-9059-8

Han, S., Wang, Y., Yang, H., Dally, W. J., Kang, J., Mao, H., & Yao, S. (2017). ESE. *Proceedings of the 2017 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays - FPGA'17*. doi:10.1145/3020078.3021745

Han, J., Kamber, M., & Pei, J. (2006). *Data mining: concepts and techniques*. Morgan kaufmann.

Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufman Publishes.

Hashem, I., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Ullah Khan, S. (2015). The rise of "big data" on cloud computing: Review and open research issues. *Information Systems*, *7*, 98–115. doi:10.1016/j.is.2014.07.006

Hinton, G., Osindero, S., & Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, *18*(7), 1527–1554. doi:10.1162/neco.2006.18.7.1527 PMID:16764513

Horticulture Statistics Division, Department of Agriculture, Cooperation & Farmers Welfare, Ministry of Agriculture & Farmers Welfare. (2017). *Horticulture Statistics At A Glance 2017*. New Delhi: Government of India Controller of Publication Indian Council of Agricultural Research. Retrieved Jan 13, 2019, from https://icar.org.in/content/horticultural_ division.

Horvath & Zelenika. (2011). *Identifcation of palaeoenvironments using clustering methods and indicator kriging, Case study from Late Miocene sandstones, the Sava depression*. Academic Press.

Huang, Y. J., & Lee, F. F. (2010). An automatic machine vision-guided grasping system for Phalaenopsis tissue culture plantlets. *Computers and Electronics in Agriculture*, *70*(1), 42–51. doi:10.1016/j.compag.2009.08.011

Hung, C., Underwood, J., Nieto, J., & Sukkarieh, S. (2015). A feature learning based approach for automated fruit yield estimation. In *Field and Service Robotics (FSR)* (pp. 485–498). Brisbane, Australia: Springer; doi:10.1007/978-3-319-07488-7_33.

Hussain, M., Gawate, S. P., Prasad, P. S., & Kamble, P. A. (2015). Smart irrigation system with three level access mechanisms. *International Conference on Computation of Power, Energy, Information and Communication (ICCPEIC)*. 10.1109/ICCPEIC.2015.7259474

ILEIA. (2000). Grassroots innovation. Unleashing the creativity of farmers. *Newsletter for Low External Input and Sustainable Agriculture*, *16*(2), 4–5.

Illingworth, J., & Kittler, J. (1987). The adaptive Hough transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-9*(5), 690–698. doi:10.1109/TPAMI.1987.4767964 PMID:21869428

Iraji, M. S., Tosinia, A., Azad, I., & Branch, A. (2011). Classification Tomatoes On Machine Vision With Fuzzy The Mamdani Inference, Adaptive Neuro Fuzzy Inference System Based (Anfis-Sugeno). *Australian Journal of Basic and Applied Sciences*, *5*(11), 846–853.

Iswari, N. M., & Wella, R. (2017). Fruitylicious: Mobile application for fruit ripeness determination based on fruit image. *2017 10th International Conference on Human System Interactions (HSI)*, 183-187.

Jain, A., & Dubes, R. (1998). *Algorithms for Clustering Data*. Englewood Cliffs, NJ: Prentice-Hall.

Jain, R., Minz, S., & Rama Subramaniam, V. (2009). Machine learning for forewarning crop diseases. *Journal of the Indian Society of Agricultural Statistics*, *63*(1), 97–107.

Jamil, N., Mohamed, A., & Abdullah, S. (2009). Automated Grading of Palm Oil Fresh Fruit Bunches (FFB) using Neuro-Fuzzy. 2009 *International Conference of Soft Computing and Pattern Recognition*, 245-249. Malacca.10.1109/SoCPaR.2009.57

Jasmine Sheela & Sivaranjani. (2015). A Brief Survey of Classification Techniques Applied To Soil Fertility Prediction. *ICETSH-2015*.

Jayarathna, H. M., & Hettige, B. (2013). AgriCom: A communication platform for agriculture sector. *2013 IEEE 8th International Conference on Industrial and Information Systems*, 439-444.

Jena, T., Rajesh, T. M., & Patil, M. (2019). Elitist TLBO for Identification and Verification of Plant Diseases. In A. Kulkarni, P. Singh, S. Satapathy, A. Husseinzadeh Kashan, & K. Tai (Eds.), *Socio-cultural Inspired Metaheuristics. Studies in Computational Intelligence* (Vol. 828). Singapore: Springer. doi:10.1007/978-981-13-6569-0_3

Jhuria, M., Kumar, A. P., & Borse, R. (2013). Image processing for smart farming: Detection of disease and fruit grading. 2013 IEEE Second International Conference on Image Information Processing (ICIIP-2013), 521-526. doi:10.1109/ICIIP.2013.6707647

Ji, B. S., & Wan, J. (2007). Artificial neutral networks for rice yield prediction in mountainous regions. *Journal of Agricultural Science*.

Jin, J., Fu, K., & Zhang, C. (2014). Traffic Sign Recognition With Hinge Loss Trained Convolutional Neural Networks. *IEEE Transactions on Intelligent Transportation Systems*, *15*(5), 1991–2000. doi:10.1109/TITS.2014.2308281

Jones, J. W. (1998). Decision support system for agrotechnology transfer: DSSAT v3. In G. Y. Tsuji, G. Hoogenboom, & P. K. Thornton (Eds.), *Understanding options for agricultural production* (pp. 157–177). Kluwer Academic Publishers. doi:10.1007/978-94-017-3624-4_8

Jorquera, H., Perez, R., Cipriano, A., & Acuna, G. (2001). Short term forecasting of air pollution episodes. In P. Zannetti (Ed.), *Environmental modeling 4*. WIT Press.

Joshi, A., & Ali, L. (2017). A detailed survey on auto irrigation system. In *2017 Conference on Emerging Devices and Smart Systems (ICEDSS)* (pp. 90–95). Mallasamudram, Tiruchengode, India: IEEE. 10.1109/ICEDSS.2017.8073665

JP, Serra & Soille, P. (1994). Mathematical Morphology and Its Applications to Image Processing (Computational Imaging and Vision). . doi:10.1007/978-94-011-1040-2

Jr, M. (2016). Automatic green fruit counting in orange trees using digital images. *Computers and Electronics in Agriculture*, *127*, 572–581. doi:10.1016/j.compag.2016.07.023

Junfitrhana, A. P., Langlangbuana, M. L., & Fatah, W. A., & Susilawati (2017). Developing potential agriculture land detector for determine suitable plant using Raspberry-Pi. *2017 International Conference on Computing, Engineering, and Design (ICCED)*, 1-4. 10.1109/CED.2017.8308116

Kader, A., & Rolle, R. (2004). *The Role Of Post-Harvest Management In Assuring The Quality And Safety Of Horticultural Produce*. Retrieved Jan 13, 2019, from http://www.fao.org /docrep/007 /y5431e/ y5431e00.htm#Contents

Kalaivani, T., Allirani, A., & Priya, P. (2011). A survey on Zigbee based wireless sensor networks in agriculture. In *3rd International Conference on Trendz in Information Sciences & Computing (TISC2011)* (pp. 85–89). Chennai, India: IEEE. 10.1109/TISC.2011.6169090

Kamble, U., Shingne, P., Kankrayane, R., Somkuwar, S., & Kamble, S. (2017). Testing of agriculture soil by digital image processing. IJSRD-International Journal for Scientific Research & Development, 5(01), 3. Retrieved from http://ijsrd.com/Article.php?manuscript=IJSRDV5I10396

Kamesh, V. (2019). Real-Time Fraud Anomaly Detection in E-banking Using Data Mining Algorithm. *South Asian Journal of Engineering and Technology*, *8*(1), 144–148.

Kamilaris, A., Gao, F., Prenafeta-Boldú, F. X., & Ali, M. I. (2016). *Agri-IoT: A Semantic Framework for Internet of Things-enabled Smart Farming Applications*. Reston, VA: Academic Press.

Kamilaris, A., Kartakoullis, A., & Prenafeta-Boldu, F. X. (2017). A review on the practice of big data analysis in agriculture. *Computers and Electronics in Agriculture*, *143*, 23–37. doi:10.1016/j.compag.2017.09.037

Kannan, P. K., & Hemalatha, K. (2012). Agro Genius: An Emergent Expert System for Querying Agricultural Clarification Using Data Mining. *International Journal of Engineering and Science, 1*(11), 34-39. Retrieved from http://rolliawati.dosen.narotama.ac.id/files/2012/03/fuzzy-toolbox-in-matlab.pdf

Kanungo, D. P., Nayak, J., Naik, B., & Behera, H. S. (2016). Hybrid clustering using elitist teaching learning-based optimization: An improved hybrid approach of PCA. *International Journal of Rough Sets and Data Analysis*, *3*(1), 1–19. doi:10.4018/IJRSDA.2016010101

Kaptan, S. S. (2002). *New Concepts in Banking*. Sarup & Sons, Edition.

Karchi & Nagesh. (2017). A Hybrid Approach for Hyper Spectral Image Segmentation Using SMLR and PSO Optimization. *Proceedings of Recent Trends in Image Processing and Pattern Recognition, Springer Nature Singapore, CCIS 709*, 1–9.

Karchi, R. P., & Munusamy, N. (2018). Exploration of Unmixing and Classification of Hyperspectral Imagery. *International Journal of Future Generation Communication and Networking*, *11*(6).

Karisiddappa, Ramegowda, & Shridhara, S. (2010). Soil characterization based on digital image analysis. In Indian Geotechnical Conference- GEOtrendz (p. 4). Retrieved from https://gndec.ac.in/~igs/ldh/conf/2010/articles/017.pdf

Karmas, A., Karantzalos, K., & Athanasiou, S. (2014). Online analysis of remote sensing data for agricultural applications. s.l. *OSGeo's European conference on free and open source software for geospatial*.

Karugia, S., Baltenweck, I., Waithaka, M., Miano, M., Nyikal, R., & Romney, D. (2004). *Perception of Technology and its Impact on Technology Uptake: The Case of Fodder Legume in Central Kenya Highlands*. The Role of Social Scientists Proceedings of the Inaugural Symposium, Nairobi, Kenya.

Kassim, M. R., Mat, I., & Harun, A. N. (2014). Wireless Sensor Network in precision agriculture application. *2014 International Conference on Computer, Information and Telecommunication Systems (CITS)*, 1-5.

Kaur, N., Singh Rddi, N., & Payyail, A. (2013). Review of Expert system based of Fuzzy logic. *International Journal of Academic Research in Computer Engineering, 2*(3). Retrieved from https://www.ijarcce.com/upload/2013/march/4-anand%20nayyar%20-20Review%20of%20Expert-c.pdf

Kavdır, I., & Guyer, D. E. (2008). Evaluation Of Different Pattern Recognition Techniques For Apple Sorting. *Biosystems Engineering*, *99*(2), 211–219. doi:10.1016/j.biosystemseng.2007.09.019

Kavianand, G., Nivas, V. M., Kiruthika, R., & Lalitha, S. (2016). *Smart drip irrigation system for sustainable agriculture. In 2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR)* (pp. 19–22). Chennai, India: IEEE; doi:10.1109/TIAR.2016.7801206

Kavyashree, D., & Rajesh, T. M. (2018, November). Analysis on Text Detection and Extraction from Complex Background Images. *Journal on Pattern Recognition*, 37-43.

Kelkar, S., Boushey, C. J., & Okos, M. (2015). A Method to Determine the Density of Foods using X-ray Imaging. *Journal of Food Engineering*, *159*, 36–41. doi:10.1016/j.jfoodeng.2015.03.012 PubMed

Keskin, M., & Sekerli, Y. E. (2016). Awareness and adoption of precision agriculture in the Cukurova region of Turkey. *Agronomy Research (Tartu)*, *14*(4), 1307–1320.

Khan, A. R., Dubey, M. K., Bisen, P. K., & Saxena, K. K. (2007). Constraints faced by farmers of Narsing Kheda village of Sihore district. *Young (up to 30 yrs.), 8*, 16.

Khanna, A., & Kaur, S. (2019). Evolution of Internet of Things (IoT) and its significant impact in the field of Precision Agriculture. *Computers and Electronics in Agriculture*, *157*, 218–231. doi:10.1016/j.compag.2018.12.039

Khazaei, J., Naghavi, M. R., Jahansouz, M. R., & Salimi-Khorshidi, G. (2008). Yield estimation and clustering of chickpea genotypes using soft computing techniques. *Agronomy Journal*, *100*(4), 1077. doi:10.2134/agronj2006.0244

Khirade, D. Sachin & Patil, A.B. (2015). Plant Disease Detection Using Image Processing. Proceedings - 1st International Conference on Computing, Communication, Control and Automation, ICCUBEA 2015. 768-771. 10.1109/ICCUBEA.2015.153

Khodaskar, K. (2015). Virtual Fruits Market – An Application for Farmer. *Fifth International Conference on Communication Systems and Network Technology*. 10.1109/CSNT.2015.235

Kim, D., Choi, H., Choi, J., Yoo, S. J., & Han, D. (2015). A novel red apple detection algorithm based on AdaBoost learning. IEIE Transactions on Smart Processing & Computing, 4(4), 265–271. doi:10.5573/IEIESPC.2015.4.4.265

Kim, G., Xing, E. P., Fei-Fei, L., & Kanade, T. (2011, November).Distributed cosegmentation via submodular optimization on anisotropic diffusion. In *2011 International Conference on Computer Vision* (pp. 169-176). IEEE.

Kimiti, J. M., Odee, D. W., & Vanlauwe, B. (2009). *Area under grain legumes cultivation and problems faced by smallholder farmers in legume production in the semi-arid eastern Kenya*. Academic Press.

Klise, K. A., & McKenna, S. A. (2006). *Water Quality Change Detection: Multivariate Algorithms. Proceedings of SPIE 6203*.

Klosgen, W., & Zytkow, J. M. (2002). *Handbook of data mining and knowledge discovery*. Oxford University Press.

Komali, A., Babu, R. V., Kumar, D. S., & Babu, K. G. (2015). An Efficient Content Based Image Retrieval System for Color and Shape Using Optimized K-Means Algorithm. *International Journal of Computer Science and Network Security*, *15*(4), 127.

Kondo, N., Ahmad, U., Monta, M., & Murase, H. (2000). Machine Vision Based Quality Evaluation OfIyokan Orange Fruit Using Neural Networks. *Computers and Electronics in Agriculture*, *29*(1-2), 135–147. doi:10.1016/S0168-1699(00)00141-1

Koprowski, R. (2014). Quantitative assessment of the impact of biomedical image acquisition on the results obtained from image analysis and processing. *Biomedical Engineering Online*, *13*(1), 93. doi:10.1186/1475-925X-13-93 PubMed

Korkut, U. B., Göktürk, Ö. B., & Yildiz, O. (2018). *Might. Detection of plant diseases by machine learning. In 2018 26th Signal Procedureing and Communications Applications Conference (SIU)*. IEEE.

Kosa, J., Bennett, F., Andre, D. & Keane, M. (1996). *Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming*. Academic Press.

Krig, S. (2014). Image Pre-Processing. In Computer Vision Metrics. Berkeley, CA: Apress. doi:10.1007/978-1-4302-5930-5_2

Krishna, P. R., & Sao, S. (2016). An improved PCA algorithm to solve profit based unit commitment problem under deregulated environment. *Procedia Technology*, *25*, 652–659. doi:10.1016/j.protcy.2016.08.157

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Communications of the ACM*, *60*(6), 84–90. doi:10.1145/3065386

Kruize, Wolfert, Scholten, Verdouw, Kassahun, & Beulens. (2016). *A Reference architecture for Farm Software Eco-systems*. Academic Press.

Kulalvaimozhi, V. P., Alex, M. G., & Peter, S. J. (2017). Image processing in agriculture. International Journal Of Advancement In Engineering Technology, Management and Applied Science (IJAETMAS), 04(03), 10. Retrieved from http://www.ijaetmas.com/wp-content/uploads/2017/04/IJ17M0314.pdf

Kumar, A. (2017, April). A Moving Object Recognition using Video Analytics. *International Journal for Scientific Research & Development*.

Kumar, R. M. (2009). Crop Selection Method to Maximize Crop Yield Rate Using Machine Learning Technique. *Energy Materials*, *1*(1).

Kumar, V., Dave, V., Nagrani, R., Chaudhary, S., & Bhise, M. (2013). Crop cultivation information system on mobile devices. *2013 IEEE Global Humanitarian Technology Conference: South Asia Satellite (GHTC-SAS)*, 196-202. 10.1109/GHTC-SAS.2013.6629915

Kumar, V., Vimal, B., Kumar, R., Kumar, R., & Kumar, M. (2014). Determination of soil pH by using digital image processing technique. *Journal of Applied and Natural Science*, *6*(1), 14–18. doi:10.31018/jans.v6i1.368

Kurtulmuş, F., & Ünal, H. (2015). Discriminating Rapeseed Varieties Using Computer Vision And Machine Learning. *Expert Systems with Applications*, *42*(4), 1880–1891. doi:10.1016/j.eswa.2014.10.003

Lashari, Ibrahim, Senan, & Taujuddin. (2018). Application of Data Mining Techniques for Medical Data Classification: A Review. *MATEC Web of Conferences, 150*, 06003.

Lasi, H., Fettke, P., Kemper, H. G., Feld, T., & Hoffmann, M. (2014). Industry 4.0. *Business & Information Systems Engineering*, *6*(4), 239–242. doi:10.100712599-014-0334-4

Lee, D. J., Chang, Y., Archibald, J. K., & Greco, C. G. (2008, August). Color Quantization And Image Analysis For Automated Fruit Quality Evaluation. In *Automation Science and Engineering, 2008.CASE 2008. IEEE International Conference on* (pp. 194-199). IEEE. 10.1109/COASE.2008.4626418

Leemans, V., & Destain, M. F. (2004). A Real-Time Grading Method Of Apples Based On Features Extracted From Defects. *Journal of Food Engineering, 61*(1), 83-89. doi: (03)00189-4 doi:10.1016/S0260-774

Leemans, V., Magein, H., & Destain, M. (2002, July). On-line Fruit Grading according to their External Quality using Machine Vision. *Biosystems Engineering*, *83*(4), 97–404.

Lee, S., & Kerschberg, L. (1998). Methodology and life cycle model for data mining and knowledge discovery in precision agriculture. *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, 3, 2882–2887.

Li, Y., Zhang, W., & Yang, H. (2012). The Application of Expert System and Neural Networks in Crop Growth Management System. *7th International Conference on Computer Science & Education (ICCSE 2012)*. Retrieved from https://ieeexplore.ieee.org/document/6295145/

Liao, Y., Deschamps, F., Loures, E. D. F. R., & Ramos, L. F. P. (2017). Past, present and future of industry 4.0-a systematic literature review and research agenda proposal. *International Journal of Production Research*, *55*(12), 3609–3629. doi:10.1080/00207543.2017.1308576

Lim, C., Kim, K.-J., & Maglio, P. P. (2018). Smart cities with big data: Reference models, challenges, and considerations. *Cities (London, England)*, *82*, 86–99. doi:10.1016/j.cities.2018.04.011

Liming, X., & Yanchao, Z. (2010, September). Automated Strawberry Grading System Based On Image Processing. *Computers and Electronics in Agriculture*, *71*(1), S32–S39. doi:10.1016/j.compag.2009.09.013

Lipton, Z., Berkowitz, J., & Elkan, C. (2015). *A Critical Review of Recurrent Neural Networks for Sequence Learning*. Retrieved from https://arxiv.org/abs/1506.00019

Li, W.-L., Xie, H., Li, Q.-D., Zhou, L.-P., & Yin, Z.-P. (2014). Section Curve Reconstruction and Mean-Camber Curve Extraction of a Point-Sampled Blade Surface. *PLoS One*, *9*(12), e115471. doi:10.1371/journal.pone.0115471 PubMed

Loevinsohn, M., Sumberg, J., & Diagne, A. (2012). Under what circumstances and conditions does adoption of technology result in increased agricultural productivity? Protocol. London: EPPI Centre, Social Science Research Unit, Institute of Education, University of London.

Lokers, Knapen, Janssen, van Randen, & Jansen. (2016). *Analysis of Big Data technologies for use in agro-environmental science.* Academic Press.

Lokers, R., Knapen, R., Janssen, S., van Randen, Y., & Jansen, J. (2016). Analysis of Big Data technologies for use in agro-environmental science. *Environmental Modelling & Software*, *84*, 494–504. doi:10.1016/j.envsoft.2016.07.017

Lorestani, A. N., Omid, M., Bagheri-Shooraki, S., Borghei, A. M., & Tabatabaeefar, A. (2006). Design And Evaluation Of A Fuzzy Logic Based Decision Support System For Grading Of Golden Delicious Apples. *International Journal of Agriculture and Biology*, *8*(4), 440–444.

Lu, C.-T., Chen, D., & Kou, Y. (2003). Algorithms for Spatial Outlier Detection. *Third IEEE International Conference on Data Mining (ICDM'03)*. 10.1109/ICDM.2003.1250986

Lu, S., Cai, Z. J., & Zhang, X. B. (2009). Forecasting Agriculture Water Consumption based on PSO and SVM. In *IEEE International Conference on Computer Science and Information Technology* (pp. 147–150). Beijing, China: IEEE.

Maharlooei, M., Sivarajan, S., Bajwa, S., Harmon, J., & Nowatzki, J. (2017). Detection of soybean aphids in a greenhouse using an image processing technique. *Computers and Electronics in Agriculture*, *132*, 63–70. doi:10.1016/j.compag.2016.11.019

Maier, H. R., & Dandy, G. C. (2000). Neural networks for the prediction and forecasting of water resources variables: A review of modelling issues and applications. *Environmental Modelling & Software*, *15*(1), 101–124. doi:10.1016/S1364-8152(99)00007-9

Maini, R., & Aggarwal, H. (2010). A Comprehensive Review of Image Enhancement Techniques. *Journal of Computers*, 2.

Majumdar, Sneha, & Shilpa. (2017). Analysis of agriculture data using data mining techniques: application of big data. *Journal of Big Data*, 4-20.

Majumdar, J., & Naraseeyappa, S., & Ankalaki. (2017). Analysis of agriculture data using data mining techniques: Application of big data S. *J. Big Data*, *4*(20).

ManikandaBabu. C. S., & Arun Pandian, M. (2016). Determination of physical and chemical characteristics of soil using digital image processing. International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE), 20(2), 5. Retrieved from http://www.ijetcse.com/wp-content/plugins/ijetcse/file/upload/docx/9ICAEECI-017-pdf.pdf

Manjula, E. (2017). Data mining technique to analyze soil nutrients based on hybrid classification. *International Journal of Advanced Research in Computer Science.*, *8*(8), 505–510. doi:10.26483/ijarcs.v8i8.4794

Manoj & Singhm. (2011). Applications of Data Mining for Intrusion Detection. *International Journal of Educational Planning and Administration*, *1*(1), 37–42.

Martin, A., Sathish, D., Balachander, C., Hariprasath, T., & Krishnamoorthi, G. (2015). Identification and counting of pests using extended region grow algorithm. 2015 2nd International Conference on Electronics and Communication Systems (ICECS), 1229-1234.

Masoud, R., & Ahmed, T. M. (2016). Using Data Mining in Telecommunication Industry: Customer's Churn Prediction Model. *Journal of Theoretical and Applied Information Technology*, *91*(2).

May, Z., & Amaran, M. (2011). Automated ripeness assessment of oil palm fruit using RGB and fuzzy logic technique. In M. Demiralp, Z. Bojkovic, & A. Repanovici (Eds.), *Proceedings of the 13th WSEAS international conference on Mathematical and computational methods in science and engineering (MACMESE'11)* (pp. 52-59). Catania, Sicily, Italy: World Scientific and Engineering Academy and Society (WSEAS).

Maynard, A. D. (2015). Navigating the fourth industrial revolution. *Nature Nanotechnology*, *10*(12), 1005–1006. doi:10.1038/nnano.2015.286 PMID:26632281

McQueen, R., & Garner, S., C.G., N.-M., & Witten, I. H. (1995). Applying machine learning to agricultural data. *Computers and Electronics in Agriculture*, *12*(1).

Megala, S., & Hemalatha, M. (2011). A Novel Datamining Approach to Determine the Vanished Agricultural Land in Tamilnadu. *International Journal of Computers and Applications*, 23. doi:10.5120/2869-3718

Mehra, T., Kumar, V., & Gupta, P. "Maturity and disease detection in tomato using computer vision," 2016 Fourth International Conference on Parallel, Distributed and Grid Computing (PDGC), Waknaghat, 2016, pp. 399-403.doi:10.1109/PDGC.2016.7913228

Mehta & Kathiriya. (2015). Survey of Data Mining Techniques in Precision Agriculture. *International Journal of Standardization Research*.

Mekala, M. S., & Viswanathan, P. (2017). A Survey: Smart agriculture IoT with cloud computing. In *2017 International conference on Microelectronic Devices, Circuits and Systems (ICMDCS)* (pp. 1–7). Vellore: IEEE. 10.1109/ICMDCS.2017.8211551

Meyer, G. E., Neto, J. C., Jones, D. D., & Hindman, T. W. (2004). Intensified fuzzy clusters for classifying plant, soil, and residue regions of interest from color images. *Computers and Electronics in Agriculture*, *42*(3), 161–180. doi:10.1016/j.compag.2003.08.002

Mhaski, R. R., Chopade, P. P., & Dale, M. P. (2015). Determination of ripeness and grading of tomato using image analysis on Raspberry Pi. *2015 Communication Control and Intelligent Systems*, 214–220.

Miller, McCarthy, & Zakzeski. (2009). A Fresh Approach to Agricultural Statistics: Data Mining and Remote Sensing. *Section on Government Statistics*, 3144-3155.

Ministry of Agriculture & Farmers Welfare. (2015). *Kisan Suvidha.* Author.

Mishra, B., Nema, S., Lambert, M., & Nema, S. "Recent technologies of leaf disease detection using image processing approach 2014; A review," *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, Coimbatore, 2017, pp. 1-5. doi: 10.1109/ICIIECS.2017.8275901

Mittal, A., & Singh, A. (2007). Microcontroller based pest management system. In *Second International Conference on Systems (ICONS'07)* (pp. 43–43). Martinique, France: IEEE. 10.1109/ICONS.2007.35

Mohamad, M. R. A., & Gombe, M. I. (2017). *E-agriculture Revisited: A Systematic Literature Review of Theories, Concept,Practices, Methods, and Future Trends*. British Academy of Management.

Mondal, A. S., Misra, I. S., & Bose, S. (2017). Building a low cost solution using wireless sensor network for agriculture application. *2017 International Conference on Innovations in Electronics, Signal Processing and Communication (IESC)*, 61-65. 10.1109/IESPC.2017.8071865

Mucherino, A., & Ruß, G. (2011). *Recent Developments in Data Mining and Agriculture*. Retrieved May 17, 2019, from http://www.antoniomucherino.it/download/papers/MucherinoRuss-dma11.pdf

Mucherino, A., & Urtubia, A. (2010). Consistent Biclustering and Applications to Agriculture. In *Proceedings of the Industrial Conference on Data Mining (ICDM10), Workshop "Data Mining in Agriculture" (DMA10)* (pp.105-113). Berlin, Germany: Academic Press.

Mucherino, A., Papajorgji, P., & Pardalos, P. (2009). *Data mining in agriculture* (Vol. 34). Springer. doi:10.1007/978-0-387-88615-2

Murakami, Y. (2014). iFarm: Development of Web-Based System of Cultivation and Cost Management for Agriculture. *2014 Eighth International Conference on Complex, Intelligent and Software Intensive Systems*, 624-627. 10.1109/CISIS.2014.89

Mustafa, N. A., Fuad, N. A., Ahmed, S. K., Abidin, A. Z., Ali, Z., & Yit, W. B. (2008). Image Processing Of An Agriculture Produce: Determination Of Size And Ripeness Of A Banana. *2008 International Symposium on Information Technology*, 1-7. doi: .463163610.1109/ITSIM.2008

Mustafa, N. B. A., Ahmed, S. K., Ali, Z., Yit, W. B., Abidin, A. A. Z., & Sharrif, Z. A. M. (2009, November). Agricultural produce sorting and grading using support vector machines and fuzzy logic. In *2009 IEEE International Conference on Signal and Image Processing Applications* (pp. 391-396). IEEE. doi: .2009.547868410.1109/ICSIPA

Mutiawani, V., Subianto, M., & Makruf, M. (2017). An Android agricultural commodity price information application by utilizing RESTful web service. *2017 International Conference on Electrical Engineering and Informatics (ICELTICs)*, 109-113. 10.1109/ICELTICS.2017.8253249

Mwangi, M., & Kariuki, S. (2015). Factors Determining Adoption of New Agricultural Technology by Smallholder Farmers in Developing Countries. *Journal of Economics and Sustainable Development*, *6*(5), 208–216.

N. Jige, Monica & R. Ratnaparkhe, Varsha. (2017). Population estimation of whitefly for cotton plant using image processing approach. 487-491. . doi:10.1109/RTEICT.2017.8256644

N. Pavithra and V. Murthy, “An Image Processing algorithm for pest detection”, pices, vol. 1, no. 3, pp. 24-26, Jul. 2017.

Nakano, K. (1997). Application Of Neural Networks To The Color Grading Of Apples. *Computers and Electronics in Agriculture*, *18*(2-3), 105–116. doi:10.1016/S0168-1699(97)00023-9

Namin, T., Esmaeilzadeh, M., Najafi, M., Brown, T. B., & Borevitz, J. O. (2017). *Deep phenotyping: Deep learning for temporal phenotype/genotype classification*. . doi:10.1101/134205

Namratha & Sharma. (2016). Educational Data Mining –Applications and Techniques. *International Journal of Latest Trends in Engineering and Technology*, *7*(2), 484–488.

Naphade, K. S. “Soil characterization using digital image processing” (1999). Theses and Dissertations. 630. https://preserve.lehigh.edu/etd/630

Narendra, V. G., & Hareesha, K. S. (2011). Cashew Kernels Classification using Colour Features. *International Journal of Machine Intelligence*, *3*(2), 52–57.

Nathan, A. (2016). *Customer segregation in banking organization using knowledge management*. Academic Press.

Navneet & Gill. (2014). Algorithm for Producing Compact Decision Trees for Enhancing Classification Accuracy in Fertilizer Recommendation of Soil. *IJCA*.

Nayak, J., Naik, B., Kanungo, D. P., & Behera, H. S. (2016). *A hybrid elicit teaching learning based optimization with fuzzy c-means (ICA-FCM) algorithm for data clustering*. Ain Shams Engineering Journal.

Nguyen, H. A., & Choi, D. (2008). Application of Data Mining to Network Intrusion Detection: Classifier Selection Model. *APNOMS. LNCS Springer*, *5297*, 399–408.

Nielson, M. (2015). *Neural Networks and Deep Learning*. Determination Press.

Ning, J., Zhang, L., Zhang, D., & Wu, C. (2010). Interactive image segmentation by maximal similarity based region merging. *Pattern Recognition*, *43*(2), 445–456. doi:10.1016/j.patcog.2009.03.004

Njenga, B. (2012). *The Application of Data Mining in the Mobile Telecommunications Industry in Kenya*. University of Nairobi.

Nuske, S., Wilshusen, K., Achar, S., Yoder, L., & Singh, S. (2014). Automated visual yield estimation in vineyards. *Journal of Field Robotics*, *31*(5), 837–860. doi:10.1002/rob.21541

O’Grady, M. J., & O’Hare, G. M. P. (2017). Modelling the smart farm. *Information Processing in Agriculture*, *4*(3), 179–187. doi:10.1016/j.inpa.2017.05.001

Obua, W. O. (2011). Machine Learning Classification Technique for Famine Prediction. *Proceedings of the World Congress on Engineering*, 2.

Omid, M., Abbasgolipour, M., Keyhani, A., & Mohtasebi, S. S. (2010). Implementation of an efficient image processing algorithm for grading raisins. *International Journal of Signal and Image Processing*, *1*(1), 31–34.

Orlando, F., Napoli, M., Dalla Marta, A., Natali, F., Mancini, M., Zanchi, C., & Orlandini, S. (2011). Growth and development responses of tobacco (Nicotianatabacum L.) to changes in physical and hydrological soil properties due to minimum tillage. *American Journal of Plant Sciences*, *2*(3), 334–344. doi:10.4236/ajps.2011.23038

P. Brinda and Dr.M.Pushparani, "Analysis of Early Leave Pest Detection", IJRITCC, May 2016.

Pandithurai, O., Aishwarya, S., Aparna, B., & Kavitha, K. (2017). Agro-tech: A digital model for monitoring soil and crops using internet of things (IOT). In *2017 Third International Conference on Science Technology Engineering & Management (ICONSTEM)* (pp. 342–346). Chennai, India: IEEE. 10.1109/ICONSTEM.2017.8261306

Parack, S., Zahid, Z., & Merchant, F. (2012). Application of data mining in educational databases for predicting academic trends and patterns. *IEEE International Conference on Technology Enhanced Education (ICTEE)*, 1-4. 10.1109/ICTEE.2012.6208617

Pardalos, P. M., Boginski, L. V., & Vazacopoulos, A. (2007). *Data mining in biomedicine*. New York: Springer. doi:10.1007/978-0-387-69319-4

Pardalos, P. M., & Hansen, P. (2008). *Data mining and mathematical programming*. American Mathematical Society. doi:10.1090/crmp/045

Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Article. *Deep Face Recognition.*, *1*(3), 1–12.

Patel, Patel, & Upadhyay. (2016). A Survey Paper of Study about Data Mining. *International Journal for Research in Computer Science*.

Patil, Naik, Beldar, & Deshpande. (2016). Smart Farming using Ardino and Data Mining. *International Conference on Computing for Sustainable Global Development*.

Patil, K. A., & Kale, N. R. (2016). A model for smart agriculture using IoT. In *2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC)* (pp. 543–545). Jalgaon, India: IEEE. 10.1109/ICGTSPICC.2016.7955360

Payne, A. B., Walsh, K. B., Subedi, P. P., & Jarvis, D. (2013). Estimation of mango crop yield using image analysis – segmentation method. *Computers and Electronics in Agriculture*, *91*, 57–64. doi:10.1016/j.compag.2012.11.009

Peng, Y., Wang, Y., Tang, J., & Li, Z. (2010). Study on rice pest insect and disease diagnosis expert system in Jiangxi province using BP Neural network. *3rd International Conference on Computer Science and information Technology*, 432-434. Retrieved from https://ieeexplore.ieee.org/document/5564128/

Peralta, Costa, Balzarini, Franco, Córdoba, & Bullock. (2015). Delineation of management zones to improve nitrogen management of wheat. *Computers and Electronics in Agriculture, 110*, 103-113.

Pivoto, D., Waquil, P. D., Talamini, E., Finocchio, C. P. S., Dalla Corte, V. F., & de Vargas Mores, G. (2018). Scientific development of smart farming technologies and their application in Brazil. *Information Processing in Agriculture*, *5*(1), 21–32. doi:10.1016/j.inpa.2017.12.002

Poultney, C., Chopra, S., & Cun, Y. L. (2006). Efficient learning of sparse representations with an energy-based model. *Advances in Neural Information Processing Systems*, 1137–1144.

Prakash, K., Moorthi, P. S., & Kumar, R. S. (2017). A study of image processing in agriculture. Int. J. Advanced Networking and Applications, 09(01), 5. Retrieved from https://www.ijana.in/papers/V9I1-4.pdf

Praveen Kumar, J., & Domnic, S. (2017, March). A new histogram based quantization method to segment and count the rosette plant leaves. In *2017 IEEE International Conference on Current Trends in Advanced Computing (ICCTAC)* (pp. 1-5). IEEE.

Praveen Kumar, J., & Domnic, S. (2018). Image Based Plant Phenotyping using Graph Based Method and Circular Hough Transform. *Journal of Information Science and Engineering*, *34*(3), 671–686.

Precision Agriculture. (2015). *What-is-precision-agriculture-and-why-is-it-important*. Retrieved May 15, 2019, from https://soilsmatter.wordpress.com/2015/02/27/what-is-precision-agriculture-and-why-is-it-important/2015

Pretty, J. (2008). Agricultural sustainability: Concepts, principles and evidence. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *363*(1491), 447–465. doi:10.1098/rstb.2007.2163 PMID:17652074

Pujari, D., Yakkundimath, R., & Byadgi, A. S. (2015). Image Processing Based Detection of Fungal Diseases in Plants. *Procedia Computer Science*, *46*, 1802–1808. doi:10.1016/j.procs.2015.02.137

Puno, John Carlo & Sybingco, Edwin & Dadios, Elmer & Valenzuela, Ira & Cuello, Joel. (2017). Determination of soil nutrients and pH level using image processing and artificial neural network. 1-6. . doi:10.1109/HNICEM.2017.8269472

Rahman, H., Sudheer Pamidimarri, D., Valarmathi, R., & Raveendran, M. (2013). *Omics: Applications in Biomedical, Agriculture and Environmental Sciences, s.l*. CRC Press.

Raichaudhuri, Rittika & Sharma, Rashmi. (2017). On Analysis of Wheat Leaf Infection by Using Image Processing. 468. 569-577. . doi:10.1007/978-981-10-1675-2_56

Rajagopalan, B., & Lall, U. (1999). A k-nearest-neighbor simulator for daily precipitation and other weather variables. *Water Resources Research*, *35*(10), 3089–3101. doi:10.1029/1999WR900028

Rajesh, T. M. & Aradhya. (2015). An Application of GMM in signature skew detection. Journal of pattern Recognition.

Rajesh, T. M., & Aradhya, V. N. (2016, September). ICA and Neural Networks for Kannada Signature Identification. *International Journal of Latest Trends in Engineering and Technology*.

Rajeswari, S., Suthendran, K., & Rajakumar, K. (2017). A smart agricultural model by integrating IoT, mobile and cloud-based big data analytics. *2017 International Conference on Intelligent Computing and Control (I2C2)*, 1-5. 10.1109/I2C2.2017.8321902

Ramesh, D., & Vishnu Vardhan, B. (2013). Data Mining Techniques and Applications to Agricultural Yield Data. *International Journal of Advanced Research in Computer and Communication Engineering*, *2*(9).

Ramesh, V., & Ramar, K. (2011). Classification of Agricultural Land Soils: A Data Mining Approach. *Agricultural Journal*, *6*(3), 82–86. doi:10.3923/aj.2011.82.86

Ramos, P. J., Prieto, F. A., Montoya, E. C., & Oliveros, C. E. (2017). Automatic fruit count on coffee branches using computer vision. Computers and Electronics in Agriculture, 137, 9e22.

Rao, R. V., & Patel, V. (2013). An improved teaching-learning-based optimization algorithm for solving unconstrained optimization problems. *Scientia Iranica*, *20*(3), 710–720.

Rao, R., & Patel, V. (2012). An elitist teaching-learning-based optimization algorithm for solving complex constrained optimization problems. *International Journal of IndustrialEngineering Computations*, *3*(4), 535–560. doi:10.5267/j.ijiec.2012.03.007

Rashmi & Munusamy. (2018). Hyperspectral Image Classification and Unmixing by using ART and SUnSPI Techniques. *International Journal of Database Theory and Application, 11*(3), 13-28.

Rashmi & Nagesh. (2013). A Review of Spectral Unmixing Algorithms in the context of Mars Dataset. *International Journal of Latest Trends in Engineering and Technology*, 55-60.

Rau, A. J., Sankar, J., Mohan, A. R., Krishna, D. D., & Mathew, J. (2017). IoT based smart irrigation system and nutrient detection with disease analysis. 2017 IEEE Region 10 Symposium (TENSYMP), 1-4.

Razavian, A. S., Azizpour, H., Sullivan, J., & Carlsson, S. (2014). CNN features off-the-shelf: an astounding baseline for recognition. *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 10.1109/CVPRW.2014.131

Reijntjes, C., & Waters-Bayer, A. (Eds.). (2001). *Farmer Innovation in Africa. A source of Inspiration for Agricultural Development*. London, UK: Earthscan Publications.

Ren, Z., & Lu, X. (2015). Design of fertilization recommendation knowledge base and application. *First International Conference on Agro- Geoinformatics (Agro-Geoinformatics)*, 1-5.

RIICE Partnership. (2014). *Remote sensing-based Information and Insurance for Crops in Emerging economies*. Retrieved May 15, 2019, from http://www.riice.org/

Riyadi, S., Ishak, A. J., Mustafa, M. M., & Hussain, A. (2008, May). Wavelet-Based Feature Extraction Technique For Fruit Shape Classification. In *Mechatronics and Its Applications, 2008.ISMA 2008. 5th International Symposium on* (pp. 1-5). IEEE. 10.1109/ISMA.2008.4648858

Rocha, A., Hauagge, D. C., Wainer, J., & Goldenstein, S. (2010). Automatic Fruit And Vegetable Classification From Images. *Computers and Electronics in Agriculture, 70*(1), 96-104. doi: .09.002 doi:10.1016/j.compag.2009

Rokunuzzaman, M., & Jayasuriya, H. P. W. (2013). Development of a low cost machine vision system for sorting of tomatoes. *Agricultural Engineering International: CIGR Journal*, *15*(1).

Roopaei, Mehdi & Rad, Paul & Raymond Choo, Kim-Kwang. (2017). Cloud of Things in Smart Agriculture: Intelligent Irrigation Monitoring by Thermal Imaging. IEEE Cloud Computing. 4. 10-15. . doi:10.1109/MCC.2017.5

Rossana, M. C. (2013). A Prediction Model Framework for Crop Yield Prediction. *Asia Pacific Industrial Engineering and Management Society Conference Proceedings*, 185.

Roy, K., Mitra, I., Kar, S., Ojha, P. K., Das, R. N., & Kabir, H. (2012). Comparative studies on some metrics for external validation of QSPR models. *Journal of Chemical Information and Modeling*, *52*(2), 396–408. doi:10.1021/ci200520g PMID:22201416

Rub & Brenning. (2010). *Data-mining in precision agriculture: Management of spatial information.* Academic Press.

Rub & Kruse. (2011). Exploratory hierarchical clustering for management zone delineation in precision agriculture. In *Advances in Data Mining. Applications and Theoretical Aspects*. Springer Berlin Heidelberg.

Ruß, G. (2009). Data Mining of Agricultural Yield Data: A Comparison of Regression Models. In Proceedings of Advances in Data Mining – Applications and Theoretical Aspects (pp. 24–37). Berlin: Springer.

Ruß, G., & Brenning, A. (2010). Data Mining in Precision Agriculture: Management of Spatial Information. Computational Intelligence for Knowledge-Based Systems Design. In *13th International Conference on Information Processing and Management of Uncertainty* (pp. 350-359). Dortmund, Germany: Academic Press.

Rußwurm, M., & Körner, M. (2018). Multi-Temporal Land Cover Classification with Sequential Recurrent Encoders. *ISPRS International Journal of Geo-Information*, *7*(4), 129. doi:10.3390/ijgi7040129

Ryu, M. (2015). Design and implementation of a connected farm for smart farming system. *ICSENS*, *1-4*. doi:10.1109/ICSENS.2015.7370624

S. ANITHA &, Dr.V.RADHA. (2010). Comparison of Image Preprocessing Techniques for Textile Texture Images. International Journal of Engineering Science and Technology. 2.

S.O. Chung, M.S.N. Kabir, Y.J. Kim,Variable Fertilizer Recommendation by Image-based Grass Growth Status,IFAC-PapersOnLine,Volume 51, Issue 17,2018,Pages 10-13,ISSN 405-8963,. doi:10.1016/j.ifacol.2018.08.053

Saad, N. (2002). *Farmer processes of experimentation and innovation – A review of literature*. Cali, Colombia: Participatory Research and Gender Analysis Program, PRGA Working Document No. 21.

Sabarish, B. A., & Karthi, R. (2018). Spatial Outlier Detection Algorithm for Trajectory-Data. *International Journal of Pure and Applied Mathematics*, *118*, 325–331.

Sak, H., & Senior, A. W. (2017). *Processing acoustic sequences using long short-term memory (LSTM) neural networks that include recurrent projection layers*. U.S. Patent No. 9,620,108.

Sankaran, S., Mishra, A., Ehsani, R., & Davis, C. (2010). A review of advanced techniques for detecting plant diseases. *Computers and Electronics in Agriculture*, *72*(1), 1–13. doi:10.1016/j.compag.2010.02.007

Sara Ghods and Vahhab Shojaeddini, "A novel automated image analysis method for counting the Population of whiteflies on leaves of crops'', J. Crop Prot., 2016.

Sawaitul, Wagh, & Chatur. (2012). Classification and Prediction of Future Weather by using Back Propagation Algorithm-An Approach. *International Journal of Emerging Technology and Advanced Engineering*, *2*(1).

Sawant, A., & Durbha. (2014). Krishisense: A Semantically Aware Web Enabled Wireless Sensornetwork System for Precision Agriculture Applications. *IEEE Geoscience and Remote Sensing Symposium*.

Say, S., Keskin, M., Sehri, M., & Sekerli, Y. (2017). *Adoption of Precision Agriculture Technologies in Developed and Developing Countries*. *International Science and Technology Conference*, Berlin, Germany.

Scharr, H., Minervini, M., Fischbach, A., & Tsaftaris, S. A. (2014, July). Annotated image datasets of rosette plants. In *Proceedings of European Conference on Computer Vision* (pp. 6-12). Zürich, Suisse: Academic Press.

Scharr, H., Minervini, M., French, A. P., Klukas, C., Kramer, D. M., Liu, X., & Yin, X. (2016). Leaf segmentation in plant phenotyping: A collation study. *Machine Vision and Applications*, *27*(4), 585–606. doi:10.100700138-015-0737-3

Schellberg, J., Hill, M. J., Gerhards, R., Rothmund, M., & Braun, M. (2008). Precision agriculture on grassland: Applications, perspectives and constraints. *European Journal of Agronomy*, *29*(2-3), 59–71. doi:10.1016/j.eja.2008.05.005

Schoeman, L., Williams, P., Plessis, A., & Manley, M. (2016). X-ray micro-computed tomography (μCT) for non-destructive characterisation of food microstructure. *Trends in Food Science & Technology*, *47*(1), 10–24. doi:10.1016/j.tifs.2015.10.016

Senanayake, R. (1991). Sustainable Agriculture: Definitions and parameters for measurement. *Journal of Sustainable Agriculture*, *1*(4), 7–28. doi:10.1300/J064v01n04_03

Seng, W. C., & Mirisaee, S. H. (2009, August). A new method for fruits recognition system. In *Electrical Engineering and Informatics, 2009.ICEEI'09.International Conference on* (Vol. 1, pp. 130-134). IEEE. doi: 10.1109/ICEEI.2009.5254804

Sereshti, H., Poursorkh, Z., Aliakbarzadeh, G., Zarre, S., & Ataolahi, S. (2017). An image analysis of TLC patterns for quality control of saffron based on soil salinity effect: A strategy for data (pre)-processing. Food Chemistry. doi: PubMed doi:10.1016/j.foodchem.2017.07.012

Serfling & Wang. (2012). *General foundations for studying masking and swamping robustness of outlier identifiers*. Academic Press.

Shahbaz, B., Ali, T., Khan, I. A., & Ahmad, M. (2010). An analysis of the problems faced by farmers in the mountains of Northwest Pakistan: Challenges for agri. extension. *Pakistan Journal of Agricultural Sciences*, *47*(4), 417–420.

Sharma, K. (2018). *Hement & Kumar, Shiv* (pp. 885–890). Soil Classification & Characterization Using Image Processing; doi:10.1109/ICCMC.2018.8488103

Sharma, L., & Mehta, N. (2012). Data Mining Techniques: A Tool for Knowledge Management System in Agriculture. *International Journal of Scientific and Technology Research*, *1*(5), 67–73.

Sharma, M., Sharma, S., & Singh, G. (2018). Performance Analysis of Statistical and Supervised Learning Techniques in Stock Data Mining. *Data (København)*, *3*(4), 54. doi:10.3390/data3040054

Sharma, M., Singh, G., & Singh, R. (2017). Stark Assessment of Lifestyle Based Human Disorders Using Data Mining Based Learning Techniques. *IRBM*, *38*(6), 305–324. doi:10.1016/j.irbm.2017.09.002

Shashi. (2009). Atmospheric Temperature Prediction using Support Vector Machines. *International Journal of Computer Theory and Engineering*, *1*(1).

Shaw, M. J., Subramaniam, C., Tan, G. W., & Welge, M. E. (2001). Knowledge management and data mining for marketing. *Decision Support Systems*, *31*(1), 127–137. doi:10.1016/S0167-9236(00)00123-8

Shekhar, Vatsavai, & Celik. (2012). Spatiotemporal data mining in the era of big spatial data: algorithms and applications. *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data*, 1-10.

Shen, W. Z., Zhang, C. L., & Chen, Z. L. (2007, August). Research on automatic counting soybean leaf aphids system based on computer vision technology. In *Proceedings of 2007 International Conference on Machine Learning and Cybernetics* (Vol. 3, pp. 1635-1638). IEEE. 10.1109/ICMLC.2007.4370408

Shit, P., Bhunia, G., & Maiti, R. (2015). Soil crack morphology analysis using image processing techniques. *Modeling Earth Systems and Environment*, *1*(4), 1–7. doi:10.1007/s40808-015-0036-z

Shivnath Ghosh, S. K. (2014). Machine Learning for Soil Fertility and Plant Nutrient Management. *International Journal on Recent and Innovation Trends in Computing*, *2*(2), 292–297.

Silwal, A., Gongal, A., & Karkee, M. (2014). Identification of red apples in field environment with over the row machine vision system. *Agricultural Engineering International: CIGR Journal*, *16*(4), 66–75.

Simonyan, K., & Zisserman, A. (2014). *Very deep convolutional networks for large-scale image recognition*. arXiv preprint arXiv:1409.1556

Singh, P.K., & Hussain, M.S. (2018). Expert System for Crop Selection in Agriculture: A Critical Review. *International Journal of Advanced Research in Computer Science, 9*(2).

Singh, V., Varsha, & Misra, A. K. (2015). Detection of unhealthy region of plant leaves using image processing and genetic algorithm. In *2015 International Conference on Advances in Computer Engineering and Applications* (pp. 1028–1032). Ghaziabad, India: IEEE. 10.1109/ICACEA.2015.7164858

Singh, A., & Sharma, T. P. (2014, July). A survey on area coverage in wireless sensor networks. In *2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)* (pp. 829-836). IEEE. 10.1109/ICCICCT.2014.6993073

Singhal, M., Verma, K., & Shukla, A. (2011). Krishi Ville — Android based solution for Indian agriculture. *2011 Fifth IEEE International Conference on Advanced Telecommunication Systems and Networks (ANTS)*, 1-5. 10.1109/ANTS.2011.6163685

Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., & Stefanovic, D. (2016). Deep neural networks based recognition of plant diseases by leaf image classification. *Computational Intelligence and Neuroscience*, *2016*, 1–11. doi:10.1155/2016/3289801 PMID:27418923

Smart Farming Challenges. (n.d.). Retrieved May 15, 2019, from https://easternpeak.com/blog/iot-in-agriculture-5-technology-use-cases-for-smart-farming-and-4-challenges-to-consider/

Snehal, S. D. (2014). Agricultural Crop Yield Prediction Using Artificial. *International Journal of Innovative Research in Electrical, Electronic*, *1*(1).

Solutions for Smart Farming. (n.d.). Retrieved January 25, 2019, from https://www.kaaproject.org/agriculture/

Somvanshi, V. K. (2006). Modeling and prediction of rainfall using artificial neural network and ARIMA techniques. *J. Ind. Geophys. Union*, *10*(2), 141–151.

Sonka, M., Hlavac, V., & Boyle, R. (1993). Image pre-processing. In *Image Processing, Analysis and Machine Vision*. Boston, MA: Springer; doi:10.1007/978-1-4899-3216-7_4.

Sonka, S. (2016). Big Data: Fueling the Next Evolution of Agricultural Innovation. *Journal of Innovation Management*, *4*(1), 114–136. doi:10.24840/2183-0606_004.001_0008

Sreekantha, D. K., Deepthi, M. B., & Puneeth, R. P. (2017). Expert Systems for Disease Diagnoses of Crops - A Study. *XIII Agricultural Science Congress-2017: Climate Smart Agriculture*.

Sriram, N., & Philip, H. (2014). *Expert System for Decision Support in Agriculture, 2014*. Retrieved from http://agritech.tnau.ac.in/pdf/14.pdf

Srisha, Ravi & Khan, Am. (2013). Morphological Operations for Image Processing: Understanding and its Applications.

Srisruthi, S., Swarna, N., Ros, G. M. S., & Elizabeth, E. (2016). Sustainable agriculture using eco-friendly and energy efficient sensor technology. In *2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)* (pp. 1442–1446). Bangalore, India: IEEE. 10.1109/RTEICT.2016.7808070

Srunitha, K & Sivasamy, Padmavathi. (2016). Performance of SVM classifier for image based soil classification. 411-415. . doi:10.1109/SCOPES.2016.7955863

Steen, K., Christiansen, P., Karstoft, H., & Jørgensen, R. (2016). Using Deep Learning to Challenge Safety Standard for Highly Autonomous Machines in Agriculture. *Journal of Imaging*, *2*(1), 6. doi:10.3390/jimaging2010006

Stein, M., Bargoti, S., & Underwood, J. (1915). 2016. Image based mango fruit detection, localisation and yield estimation using multiple view geometry. *Sensors (Basel)*, *16*, 1–25.

Stockle, C. O., Martin, S. A., & Campbell, G. S. (1994). CropSyst, a cropping systems model: Independent Component Analysis. *Journal of Machine Learning Research*, *5*, 239–253.

Studio encoding parameters of digital television for standard 4:3 and wide screen 16:9 aspect ratios, International Telecommunication Union Recommendation BT.601-7, 2011.

Su. (2011). Statistical Geo-computing: Spatial Outlier Detection in Precision Agriculture. Academic Press.

Sundmaeker, H., Verdouw, C., Wolfert, S., & Pérez Freire, L. (2016). Internet of food and farm 2020. In O. Vermesan & P. Friess (Eds.), *Digitising the Industry - Internet of Things Connecting Physical, Digital and Virtual Worlds* (pp. 129–151). River Publishers.

Sun, Y., Cheng, H., Cheng, Q., Zhou, H., Li, M., Fan, Y., ... Lammers, S. (2017). A smart-vision algorithm for counting whiteflies and thrips on sticky traps using two-dimensional Fourier transform spectrum. *Biosystems Engineering*, *153*, 82–88. doi:10.1016/j.biosystemseng.2016.11.001

Sushma R. Huddar and SwarnaGowri, Keerthana K., Vasanthi S and Sudhir Rao Rupanagudi, "Novel Algorithm for Segmentation and Automatic Identification of Pests on Plants using Image Processing'', ICCCNT'12, July 2012.

Taechatanasat & Armstrong. (2014). Decision Support System Data for Farmer Decision Making. *Proceedings of Asian Federation for Information Technology in Agriculture*, 472-486.

Tajane, V., & Janwe, N. J. (2014). Medicinal Plants Disease Identification Using Canny Edge Detection Algorithm, Analysis and CBIR. *International Journal of Advance Research in Computer Science and Software Engineering*, *4*(6), 530–536.

Tang, X., Liu, M., Zhao, H., & Tao, W. (2009, October). Leaf extraction from complicated background. In *Proceedings of 2nd International Congress on Image and Signal Processing, 2009.CISP'09* (pp. 1-5). IEEE.

Tatte, M.K., & Nichat, M.K. (2013). Enhancement in Agro Expert System for Rice Crop. *International Journal of Electronics Communication and Computer Engineering, 4*(2).

Thakura, M., Olafssonb, S., Leeb, J. S., & Hurburgha, C. R. (2010). Data Mining for Recognizing Patterns in Foodborne Disease Outbreaks. *Journal of Food Engineering*, *97*(2), 213–227. doi:10.1016/j.jfoodeng.2009.10.012

Thankachan, S., & Kirubakaran, D. (2014). A Survey Conducted on E-Agriculture with Indian Farmers. *International Journal of Computer Science and Mobile Computing*.

Tichkule, S. K., & Gawali, D. H. 2016, November. Plant diseases detection using image procedureing techniques. In *Green Engineering and Technologies (IC-GET), 2016 Online International Conference on* (pp. 1-6). IEEE.

Toreti, A., Maiorano, A., De Sanctis, G., Webber, H., Ruane, A. C., Fumagalli, D., ... Zampieri, M. (2019). Using re-analysis in crop monitoring and forecasting systems. *Agricultural Systems*, *168*, 144–153. doi:10.1016/j.agsy.2018.07.001 PMID:30774182

Tripathi, S., Srinivas, V. V., & Nanjundiah, R. S. (2006). Downscaling of precipitation for climate change scenarios: A support vector machine approach. *Journal of Hydrology (Amsterdam)*, *330*(3-4), 621–640. doi:10.1016/j.jhydrol.2006.04.030

Tripathy, A. K., Adinarayana, J., & Sudharsan, D. (2009). Geospatial Data Mining for Agriculture Pest Management - a Framework. *17th International Conference on Geoinformatics*, 1-6. 10.1109/GEOINFORMATICS.2009.5293296

Tsai, C.-M., & Yeh, Z.-M. (2008). Contrast enhancement by automatic and parameter-free piecewise linear transformation for color images. Consumer Electronics. *IEEE Transactions on.*, *54*, 213–219. doi:10.1109/TCE.2008.4560077

Tyagi, A. C. (2016). Towards a Second Green Revolution. *Irrigation and Drainage*, *65*(4), 388–389. doi:10.1002/ird.2076

Ubbens, J. R., & Stavness, I. (2017). Deep plant phenomics: A deep learning platform for complex plant phenotyping tasks. *Frontiers in Plant Science*, *8*, 1190. doi:10.3389/fpls.2017.01190 PMID:28736569

Unay, D., & Gosselin, B. (2005b). Artificial Neural Network-Based Segmentation And Apple Grading By Machine Vision. In *Image Processing, 2005.ICIP 2005.IEEE International Conference on* (Vol. 2). IEEE.10.1109/ICIP.2005.1530134

Unay, D., & Gosselin, B. (2005a). Thresholding-Based Segmentation And Apple Grading By Machine Vision. *Signal Processing Conference, 2005 13th European*, 1-4.

Unay, D., Gosselin, B., Kleynen, O., Leemans, V., Destain, M. F., & Debeir, O. (2011). Automatic Grading of Bicolored Apples by Multispectral Machine Vision. *Computers and Electronics in Agriculture*, *75*(1), 204–212. doi:10.1016/j.compag.2010.11.006

United Nations. (2017). *World population projected to reach 9.8 billion in 2050, and 11.2 billion in 2100*. Retrieved from https://www.un.org/development/desa/en/news/population/world-population-prospects-2017.html

Valliammai, N., & Geethaiakshmi, S. N. "Multiple noise reduction using hybrid method for leaf recognition," 2012 International Conference on Devices, Circuits and Systems (ICDCS), 15-16 March 2012, Coimbatore, India. doi:10.1109/ICDCSyst.2012.6188689

Verheyen, K., Adriaens, D., Hermy, M., & Deckers, S. (2001). High-resolution continuous soil classification using morphological soil profile descriptions. *Geoderma*, *101*(3-4), 31–48. doi:10.1016/S0016-7061(00)00088-4

Vibhute, Anup &, S.K.Bodhe. 2012. Application of Image Processing in Agriculture:Survey, A. (●●●).. . *International Journal of Computers and Applications*, *52*, 34–40. doi:10.5120/8176-1495

Vibhute, A., & Bodhe, S. K. (2012). Article: Applications of Image Processing in Agriculture: A Survey. *International Journal of Computers and Applications*, *52*, 34–40. doi:10.5120/8176-1495

Vijaykumar, A., & Rajinikanth, T. (2013). Estimation of the Influence of Fertilizer Nutrients Consumption on the Wheat Crop yield in India- a Data mining Approach. *International Journal of Engineering and Advanced Technology*, *3*(2), 316–320.

Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P. A. (2008). Extracting and composing robust features with denoising autoencoders. *Proc. 25th Int.Conf. on Machine Learning*, 1096–1103.

Wang, Q.I. & Nuske, Stephen & Bergerman, Marcel & Singh, Sanjiv. (2012). Automated Crop Yield Estimation for Apple Orchards. Experimental Robotics: The 13th International Symposium on Experimental Robotics. 88. . doi:10.1007/978-3-319-00065-7_50

Wang, Xia, & Wang. (2016). A *Spectral Clustering Basedoutlier detection technique*. Springer International Publishing.

Wang, Z. & Mujib, A. B. M. (2017). The Weather Forecast Using Data Mining Research Based on Cloud Computing. *Journal of Physics: Conference Series*, *910*, 012020. doi:10.1088/1742-6596/910/1/012020

Weber, R. H., & Weber, R. (2010). *Internet of Things*. New York: Springer. doi:10.1007/978-3-642-11710-7

Wechsler, H. (Ed.). (1992). Neural Networks for Perception. Academic Press, Inc.

Weiss, G. (2009). *Data Mining in the Telecommunications Industry*. Academic Press.

Wolfert, Ge, Verdouw, & Bogaardt. (2017). Big Data in Smart Farming–A review. *Agricultural System*, (153), 69-80.

Wolfert, J., Sørensen, G., & Goense, D. (2014). A Future Internet Collaboration Platform for Safe and Healthy Food from Farm to Fork. In *Global Conference (SRII)*, (pp. 266–273). IEEE. 10.1109/SRII.2014.47

Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M.-J. (2017). Big Data in Smart Farming – A review. *Agricultural Systems*, *153*, 69–80. doi:10.1016/j.agsy.2017.01.023

Wu, H., Zhou, Y., Luo, Q., & Basset, M. A. (2016). Training Feedforward Neural Networks Using Symbiotic Organisms Search Algorithm. *Computational Intelligence and Neuroscience*, *2016*, 1–14. doi:10.1155/2016/9063065 PMID:28105044

Xiaobo, Z., Jiewen, Z., & Yanxiao, L. (2007, November). Apple Color Grading Based On Organization Feature Parameters. *Pattern Recognition Letters*, *28*(15), 2046–2053. doi:10.1016/j.patrec.2007.06.001

Xia, X. L., & Han, R. Y. (2017). Research on classifications of mobile networking device users based on BP neural networks. *2nd International Conference on Mechatronics and Information Technology (ICMIT 2017)*, 104–109.

Xu, R., & Wunsch, D. (2005). Survey of clustering algorithms. *Neural Networks. IEEE Transactions on*, *16*(3), 645–678. PMID:15940994

Yadav, R. K., & Khatri, R. (2016). A Weather Forecasting Model using the Data Mining Technique. *International Journal of Computers and Applications*, *139*(14), 4–12. doi:10.5120/ijca2016908900

Yamamoto, K., Ninomiya, S., Kimura, Y., Hashimoto, A., Yoshioka, Y., & Kameoka, T. (2015, January). Strawberry Cultivar Identification And Quality Evaluation On The Basis Of Multiple Fruit Appearance Features. *Computers and Electronics in Agriculture*, *110*, 233–240. doi:10.1016/j.compag.2014.11.018

Yang, L., Dickinson, J., Wu, Q. J., & Lang, S. (2007, December). A Fruit Recognition Method For Automatic Harvesting. In *2007 14th International Conference on Mechatronics and Machine Vision in Practice* (pp. 152-157). IEEE. 10.1109/MMVIP.2007.4430734

Yin, X., Liu, X., Chen, J., & Kramer, D. M. (2014, October).Multi-leaf tracking from fluorescence plant videos.I n *Proceedings of 2014 IEEE International Conference on Image Processing (ICIP)*, (pp. 408-412). IEEE. 10.1109/ICIP.2014.7025081

Yong, W., Shuaishuai, L., Li, L., Minzan, L., Ming, L., Arvanitis, K. G., ... Sigrimis, N. (2018). Smart Sensors from Ground to Cloud and Web Intelligence. *IFAC-PapersOnLine*, *51*(17), 31–38. doi:10.1016/j.ifacol.2018.08.057

Zaki, M. J. (1999). Parallel and distributed association mining: A survey. *IEEE Concurrency*, *7*(4), 14–25. doi:10.1109/4434.806975

Zeiler, M., & Fergus, R. (2014). *Visualizing and Understanding Convolutional Networks*. Retrieved form https://arxiv.org/pdf/1311.2901.pdf

Zhang, Jiang, Qiu, Qiu, Wang, & Zhu. (2017). An improved method of delineating rectangular management zones using a semivariogram-based technique. Computers and Electronics in Agriculture, 121, 74-83.

Zhang, L., Gu, Z., & Li, H. (2013, September). SDSP: A novel saliency detection method by combining simple priors. In *Proceedings of 2013 20th IEEE International Conference on Image Processing (ICIP)*, (pp. 171-175). IEEE. 10.1109/ICIP.2013.6738036

Zhang, N., Wang, M., & Wang, N. (2002). Precision agriculture—A worldwide overview. *Computers and Electronics in Agriculture*, *36*(2–3), 113–132. doi:10.1016/S0168-1699(02)00096-0

Zhang, S., Zhang, C., & Yang, Q. (2003). Data preparation for data mining. *Applied Artificial Intelligence: An International Journal*, *17*(5-6), 375–381. doi:10.1080/713827180

Zulkifli Bin Husin., Abdul Hallis Bin Abdul Aziz, Ali Yeon Bin Md Shakaff Rohani Binti S Mohamed Farook, "Feasibility Study on Plant Chili Disease Detection Using Image Processing Techniques", 2012 Third International Conference on Intelligent Systems Modelling and Simulation.

About the Contributors

N. Pradeep has15 years of teaching experience in CS and E dept., Bapuji Institute of Engineering and Technology, Davangere, Karnataka, India. His experience includes 7 years research experience. He has worked at various levels namely Lecturer, Sr. Lecturer, Assistant Professor. Presently working as Associate Professor and PG Coordinator, CSE dept and Placement officer, BIET, Davangere. He has published 6 research papers in various peer-reviewed International Journals and 13 research papers in various International Conferences. He has guided more than 20 students at PG level and 30 batches at UG level.

Sandeep Kautish is a seasoned academician with over 13 Years Quality Experience in Academics since 2004. He holds PhD, MTech, MCA, MSc, PGDIM and specialized in Computer Science domain (Data mining specialization with social network analytics topics). Dr Sandeep is internationally recognized PhD Supervisor/Thesis Examiner in Computer Science. One PhD has already been awarded under my supervision from Bharathiar University Coimbatore and three in process. He is member of Editorial boards of reputed journals i.e. Image Mining (Inderscience Publications), IGI Global Journal of Decision Support Systems, Australasian Journal of Information Systems (ESCI and SCOPUS Indexed) and more than 20 journals of national and international repute. He has been nominated as Technical Program Committee member of 04 International Conferences at Bangkok, Paris, Singapore and others. Active member of CSI, IEEE, CSTA, SCIEI, IACSIT, IABE, IAENG and WASET and many more. Teaching/Academic Research Interests includes Data Mining & Social Network Analytics, Business Intelligence, Knowledge and Information Engineering.

C. R. Nirmala works as professor and head, department of Computer Science and Engineering of BIET, Davangere has got 24 years of teaching experience in teaching, mentoring and guiding the students of UG, PG and research programs of CS & E department. She received her doctorate degree from Visvesvaray Technological University, Belgaum and carried out her research at BIET. She has completed her master's degree from National Institute of Technology Karnataka, Surthkal in the year 2001 with first class and specialization in Computer Engineering and B.E in Electronics and Communication from Bapuji Institute of Engineering and Technology, Davangere in the year 1992 with first class. She started her career at Bharath Electronics Ltd, Bangalore as Trainee Engineer, then joined Bapuji Institute of Engineering and Technology as a lecturer in the department of E & C from 1995 and in CS and E dept from 2001. Acting as a single point of contact for Infosys Campus Connect, IBM Academic Activities, VMWARE IT Academy and ICT Academy India activities. IBM India Ltd, awarded her with prestigious "Drona" award, for 5 times for mentoring award-winning student projects in the annual national level software development contest- “The Great Mind Challenge” and for showing outstanding performance in the IBM academic initiative Activities in the institute. Infosys has felicitated with certificate of appreciation for outstanding performance in the Infosys campus Connect activities. Wipro Mission 10X team has recognized “Champion Trainer”. And from the institute she obtained the best teacher award for 10 times. Published and presented more than 30 papers in the International Conferences and journals like IEEE transaction, ACM and Other referred journals. 2 research scholars have been awarded with PhD Degrees and 5 research scholars are carrying out research under her guidance. 1 Indian patent has been filed in the area of wireless sensor networks and waiting for publishing. I1 book chapter is published in IGI Press on Wireless Sensor Networks.

* * *

M. Anandhavalli completed Ph.D. from Sikkim Manipal University of Computer Science Engineering, India. Her areas of interest are data mining and soft computing. She has published more than 20 research publications in highly reputed international journals and with the h index of 5, and with the total citations around 170. She had attended and presented in more than 20 national & international conferences. She served as a principal investigator for the funded project from All India Council for Technical Education (AICTE) and served as a manuscript reviewer for renowned international journals for past 8 years, and Ph.D. thesis reviewer for several universities. She has been invited to give guest lectures and serving as editorial board member for a few journals and conferences. Currently, she is an Associate Professor, at the College of Computer Science, King Khalid University, Abha, Kingdom of Saudi Arabia.

P. Shanthi Bala working as an Assistant Professor in the Department of Computer Science, School of Engineering & technology, Pondicherry University, Puducherry. Her area of Interest includes Cyber Security, Distributed Computing Systems, Knowledge Engineering, Networks, Ontology.

Kavyashree Dalawai working as Assistant Professor in the Department of Computer Science & Engineering in Dayananda Sagar university, has completed her M.Tech in Computer Science and Engineering from P.A college of Engineering, Mangalore under Visvesvaraya Technological University and B.E in Information Science & Engineering from SECAB Institute of Engineering and technology, Vijayapur under Visvesvaraya Technological University. She has worked in various reputed institutions with a total teaching experience of 4 years. Her research publication includes 1 international journal. Her research areas are Video Analytics, Pattern Recognition and Artificial Intelligence.

Arun Kumar G. H. is currently working as Assistant Professor in the Dept. of Computer Science and Engineering with 6 years of teaching experience and 3 years of research experience in domain of Data Analytics and Precision Agriculture. He has monitored and guided around 10 Under Graduate projects. His research areas of interest include Precision Agriculture, Data Analytics, Data Mining and Machine Learning. He has published 2 research articles in International conference.

Sanjeevakumar M. Hatture received the Bachelor’s Degree in Electronics and Communication Engineering from Karnataka University, Dharwad, Karnataka, India, and the Master Degree in Computer Science and Engineering from Visvesvaraya Technological University, Belagavi, Karnataka, India, Ph.D Degree in Computer and Information Sciences from Visvesvaraya Technological University, Belagavi, Karnataka, India. His research interests include biometrics, image processing, pattern recognition, soft-computing, network security, Internet of Things. He is reviewer of many IEEE Conferences, IEEE Transactions, IET Biometrics etc. He is life member of professional bodies like IEEE, IEI, ISTE and IRED.

J. Praveen Kumar received the B.Tech. (CSE) degree from Amrita University, Coimbatore, India in 2009 and M.E (Software Engineering) degree from Anna University, Tiruchirappalli, India in 2011. He is currently pursuing Ph.D. degree at National Institute of Technology, Tiruchirappalli, India. His research interests include Image processing and Software testing.

Shiv Kumar is a junior at Netaji Subhas University of Technology, pursuing a bachelor's degree in Information Technology. His research interests are Machine Learning, Deep Learning, Computer Vision and their applications in different domains.

Roopa G. M. is currently working as Assistant Professor in the Dept. of Computer Science and Engineering with 10 years of teaching experience, 2 years of Industry experience in TCS and 5 years of research experience. Her research areas of interest include Data Mining, Artificial Intelligence, IoT, Data Science and Machine Learning. She has guided 30 UG and 10 PG Projects. Her research papers are published in peer reviewed international journals. She has received "Best Session Paper" in International Conference held at Dubai and National Conference.

Susen Naik received the Bachelor's Degree in Computer Science and Engineering from Visvesvaraya Technological University, Belagavi, Karnataka, India, and the Master Degree in Computer Science and Engineering from Visvesvaraya Technological University, Belagavi, Karnataka State, India. Her research interests include Andriod application, biometrics, image processing, pattern recognition, soft-computing and network security.

Naveen Kumar K. R. is currently Currently working as Assistant Professor in the department of Computer Science & Engineering, Bapuji Institute of Engineering and Technology (BIET), Davangere. He is having six years of teaching experience and two years of Industrial experience as a Software Developer at CloudByte Tech Pvt Ltd, Bengaluru. In his academic carrier, he has taught various subjects like Data Mining, Machine Learning, Big data Analytics, Data Structures etc. for under graduate students. His prime areas of research interest include Data Mining, Precision Agriculture, Natural Language Processing and Data Analytics. He is passion in R, Python, Hadoop, and JavaScript languages. He has published his research articles in 02 Journal and International conference.

Arun Kumar R. is working in the capacity of Assistant Professor in Maratha Mandal Engineering College affiliated to Visvesvaraya Technological university (VTU), Belagavi, Karnataka, India. He completed his Ph. D from VTU with KLS Gogte Institute of Technology, Belagavi as the Research Center. His areas of interest include image processing and machine learning.

Vijay S. Rajpurohit is currently working in the capacity of Professor in KLS Gogte Institute of Technology, Belagavi. He is currently guiding research scholars in Visvesvaraya Technological University, Belagavi. His ares of interest include robotics, digital image processing, machine vision and artificial intelligence.

Domnic S. received the B.Sc. and M.C.A degrees from Bharathidasan University, Tiruchirappalli, India, in 1998 and 2001, respectively, and the Ph.D. degree from Gandhigram Rural University, Dindigul, India, in 2008. He is currently an Associate Professor with the Department of Computer Applications, National Institute of Technology at Tiruchirappalli, Tiruchirappalli. His current research interests are data compression, image/video processing, and Deep Learning.

Swati Sah joined Patan College for Professional Studies (PCPS), Nepal as HoD of Computer Science Dept. in May 2018. She has an excellent experience of 6+ years in research and teaching. Her interest area of research are Artificial Intelligence, Machine Learning, Project Management etc. She was associated with Bharati Vidyapeeth Deemed University,Pune during 2010 to 2017. She got awarded her PhD degree in Computer Application from Bharati Vidyapeeth University, Pune, India, in 2016. The topic of her PhD research was "Evaluating Optimal Clustering Techniques for Efficient Storagre Retrieval Methods in Large DataBase Using Soft Computing Techniques". She did her MCA from UP Technical University, Lucknow, India. She has also been awarded the Master degree M.P.M from Birmingham City University, United Kingdom U. K. She has also been proactively involved with professional associations and professional bodies. She has attended and organized various FDP/Seminar/workshop/Conferences. She has published her research papers in National and International journals.

Deepak Kumar Sharma received the B.Tech. in Computer Science & Engineering from G. G. S. Indraprastaha University, M. E. in Computer Technology & Applications, and Ph.D in Computer Engineering from University of Delhi, India. He is presently working as Assistant Professor in the Division of Information Technology, Netaji Subhas Institute of Technology, Dwarka, New Delhi, India. His research interests include opportunistic networks, wireless ad hoc networks, sensor networks and network security.

D. K. Sreekantha is serving Professor in Dept. of Computer Science and Engineering at NMAM Institute of Technology, Nitte, Karnataka, India. He has published about 20 papers in international journals and presented 15 papers international and national conferences in India and Paris, France. He is also a reviewer for few international journals. His research interests are Soft computing, Expert system design, IOT and Risk management.

Rajesh T. M. has completed his Ph.D. in Computer Science & Engineering from Jain University, Bangalore. His research area includes Image Processing and Pattern Recognition. M.Tech. from Visvesvaraya Technological University, Belgaum and is currently working as an Assistant Professor in Department of Computer Science & Engineering. He has 2 years of industrial experience at IGATE Global Solutions Ltd, Bangalore, one year of research experience and 3+ years of teaching experience. He is guiding 3 Ph.D. scholars and guided 5 plus M. Tech students. He has published 1 International Textbook, 6 International Journals, 1 International Conference and 3 National Conference to his credit. He worked in Database Management System project for production Support for a client in Australia. His areas of interest are Video Analytics, DBMS, Image Processing, Pattern Recognition, and Software Estimation.

Aspira S. Tripathy is a final year undergraduate in the Information Technology branch at Netaji Subhas University of Technology (Formerly known as NSIT, University of Delhi). Her research interests include Computer vision, Deep learning, Machine learning and image processing. She has authored a paper on image forensics previously and is currently working under the domain of biomedical image analysis.

Agrima Yadav is a senior at Netaji Subhas University of Technology, pursuing a bachelor's degree in Information Technology. Her interests lie in Policy making, Machine Learning, Deep Learning and their applications in Natural Language Processing.

Rana Khudhair Abbas Ahmed Zwayen is a Professor and moreover an "Academician by Choice" since Oct 2001. She has had an outstanding meritorious academic record and is a work product of prestigious systems throughout Alrafidain University College, Iraq, Baghdad. She obtained M.Sc. in Computer Science and Ph. D in progress. Additionally, she has attended One Month Full Time Faculty Training Program (FTP) from Alrafidain University College, Iraq, Baghdad in 2002 in which she topped in the batch. She has published about 18 Research Papers in reputed Journals. She has authored three books. She is Editor, Reviewer, Technical Committee Member, for 40 Plus Technical Journals and Conferences at International Levels. Rana Khudhair Abbas Ahmed Zwayen has an overall Excellence in Academic, International Academic Coordination, Research, Infrastructure and Systemic Development of organization. Her current focus is to enhance the Quality of Technical Education through strategic Quality Initiatives. Now also working as an Academic Coordinator in the Civil Rights Foundation for Culture and Media (N.G.O.) (Iraq and Lebanon). Specialties: Academic Administration, Teaching & Research, Academic Coordination. With 17 years working, Rana Khudhair Abbas Ahmed Zwayen is having a proven track record of handling tough and time bound assignments efficiently under tight delivery deadlines. Systematic and organized approach, quick adaptability, comprehensive problem detection/solving ability and logical and analytical approach are biggest assets in her account.

Index