Research Insights

Electrocardiogram Signal Classification and Machine Learning

Emerging Research and Opportunities

Sara Moein

Electrocardiogram Signal Classification and Machine Learning:

Emerging Research and Opportunities

Sara Moein
Mount Sinai School of Medicine, USA

A volume in the Advances in Medical Technologies and Clinical Practice (AMTCP) Book Series

Published in the United States of America by
IGI Global
Medical Information 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: Moein, Sara, 1983- author.
Title: Electrocardiogram signal classification and machine learning : emerging research and opportunities / by Sara Moein.
Description: Hershey PA : Medical Information Science Reference, [2018] | Includes bibliographical references.
Identifiers: LCCN 2017048281| ISBN 9781522555803 (hardcover) | ISBN 9781522555810 (ebook)
Subjects: | MESH: Electrocardiography--methods | Signal Processing, Computer-Assisted | Machine Learning | Heart Diseases--diagnosis | Pattern Recognition, Automated
Classification: LCC RC683.5.E5 | NLM WG 140 | DDC 616.1/207547--dc23 LC record available at https://lccn.loc.gov/2017048281

This book is published in the IGI Global book series Advances in Medical Technologies and Clinical Practice (AMTCP) (ISSN: 2327-9354; eISSN: 2327-9370)

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

Advances in Medical Technologies and Clinical Practice (AMTCP) Book Series

ISSN:2327-9354
EISSN:2327-9370

Editor-in-Chief: Srikanta Patnaik, SOA University, India & Priti Das, S.C.B. Medical College, India

MISSION

Medical technological innovation continues to provide avenues of research for faster and safer diagnosis and treatments for patients. Practitioners must stay up to date with these latest advancements to provide the best care for nursing and clinical practices.

The **Advances in Medical Technologies and Clinical Practice (AMTCP) Book Series** brings together the most recent research on the latest technology used in areas of nursing informatics, clinical technology, biomedicine, diagnostic technologies, and more. Researchers, students, and practitioners in this field will benefit from this fundamental coverage on the use of technology in clinical practices.

COVERAGE

- Medical informatics
- Biometrics
- Nursing Informatics
- Medical Imaging
- Clinical Nutrition
- Clinical Data Mining
- Nutrition
- Telemedicine
- Biomedical Applications
- Neural Engineering

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 Medical Technologies and Clinical Practice (AMTCP) Book Series (ISSN 2327-9354) 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-medical-technologies-clinical-practice/73682. Postmaster: Send all address changes to above address.

Titles in this Series

For a list of additional titles in this series, please visit:
https://www.igi-global.com/book-series/advances-medical-technologies-clinical-practice/73682

Biomedical Signal and Image Processing in Patient Care
Maheshkumar H. Kolekar (Indian Institute of Technology Patna, India) and Vinod Kumar (Indian Institute of Technology Roorkee,India)
Medical Information Science Reference • ©2018 • 312pp • H/C(ISBN: 9781522528296) • US $265.00

Next-Generation Mobile and Pervasive Healthcare Solutions
Jose Machado (University of Minho, Portugal) António Abelha (University of Minho, Portugal) Manuel Filipe Santos (University of Minho, Portugal) and Filipe Portela (University of Minho, Portugal)
Medical Information Science Reference • ©2018 • 286pp • H/C(ISBN: 9781522528517) • US $245.00

Innovative Research in Thermal Imaging for Biology and Medicine
Ricardo Vardasca (University of Porto, Portugal) and Joaquim Gabriel Mendes (University of Porto, Portugal)
Medical Information Science Reference • ©2017 • 340pp • H/C(ISBN: 9781522520726) • US $210.00

Internet of Things and Advanced Application in Healthcare
Catarina I. Reis (Polytechnic Institute of Leiria, Portugal) and Marisa da Silva Maximiano (Polytechnic Institute of Leiria, Portugal)
Medical Information Science Reference • ©2017 • 349pp • H/C (ISBN: 9781522518204) • US $210.00

Integrating Biologically-Inspired Nanotechnology into Medical Practice
B.K. Nayak (K.M. Centre for Post Graduate Studies, India) Anima Nanda (Sathyabama University, India) and M. Amin Bhat (Sathyabama University, India)
Medical Information Science Reference • ©2017 • 394pp • H/C (ISBN: 9781522506102) • US $190.00

Recent Advances in Drug Delivery Technology
Raj K. Keservani (Rajiv Gandhi Proudyogiki Vishwavidyalaya, India) Anil K. Sharma (Delhi Institute of Pharmaceutical Sciences and Research, India) and Rajesh Kumar Kesharwani (National Institute of Technology, Warangal, India)
Medical Information Science Reference • ©2017 • 510pp • H/C (ISBN: 9781522507543) • US $215.00

For an entire list of titles in this series, please visit:
https://www.igi-global.com/book-series/advances-medical-technologies-clinical-practice/73682

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

To the love of my life, Amin
and
My dearest parents

Table of Contents

Preface

Applying artificial intelligence (AI) for increasing the reliability for medical decision-making has been studied since some years ago and many researchers have studied in this area. Artificial intelligence has no point to be applied in a field such as medical diagnosis, unless it provides acceptable conditions for solving the problems in this area. The improvements include solving the problem with fewer resources in terms of time, money and people. In the present book, the applications of artificial intelligence to attain solutions for problems that would be difficult for other techniques to be achieved are investigated. In medical science and biology, there are research questions that cannot be easily solved with the traditional methods. Some search problems are straightforward and readily solved by existing methods. Some searches, however, cannot be solved by either statistical-mathematical methods, or by all-encompassing (so called 'brute force') informatics algorithms.

Because of the importance of using intelligent methods for increasing the accuracy of medical decision-making systems, using artificial intelligence for medical diagnosis has become a major task. In 1984, Clancey and Shortliffe provided the following definition: 'medical artificial intelligence is primarily concerned with the construction of AI programs that perform diagnosis and make therapy recommendations'. This is important about medical artificial intelligence that it is definitely an interdisciplinary field. Researchers in such interdisciplinary field are required to have the knowledge in both sides to be able to answer the questions. In this way, they can find a way to do a valid research. This work has focused on using artificial intelligence is heart disorders diagnosis.

Heart is an important organ in the body, for pumping the blood throughout the human body. The heart generates signals, which hide information in their structure. This signal information is very useful to physicians for heart disorder detection, but is not easily perceived by them. Therefore, it is necessary

that signals be decoded to be useful for interpretation by specialists. The interpretation process is sometimes easy when it only involves visual inspection of the signal. However, there are signals whose complexity is high due to the heart disorders, which affect the decoded form of signals. Automated methods can assist specialists in accurate diagnosis and therefore, automated heart signal processing becomes a reliable tool for finding important information that is hidden in the signal (Adlam & Hampton, 1997). Electrocardiogram (ECG) is an important biomedical tool for the diagnosis of heart disorders. Recent studies have worked a lot on designing automatic diagnosis systems to help physicians. However, automatic study of ECG patterns and heart rate variability is difficult due to the large variation in the morphologies of heart waveforms, not only of different patients or patient groups but also within the same patient. One of the advanced tools in automatic medical diagnosis is artificial neural network. A generic artificial neural network can be defined as a computational system consisting of a set of highly interconnected processing elements, called neurons, which process information as a response to external stimuli. An artificial neuron is a simplistic representation that emulates the signal integration and threshold firing behaviour of biological neurons by means of mathematical equations. Like their biological counterpart, artificial neurons are bound together by connections that determine the flow of information between peer neurons. Stimuli are transmitted from one processing element to another via synapses or interconnections, which can be excitatory or inhibitory. If the input to a neuron is excitatory, it is more likely that this neuron will transmit an excitatory signal to the other neurons connected to it. Whereas an inhibitory input will most likely be propagated as inhibitory.

Artificial neural networks are computational paradigms based on mathematical models that unlike traditional computing have a structure and operation that resembles that of the mammal brain. Artificial neural networks or neural networks for short, are also called connectionist systems, parallel distributed systems or adaptive systems, because they are composed by a series of interconnected processing elements that operate in parallel. Neural networks lack centralized control in the classical sense, since all the interconnected processing elements change or "adapt" simultaneously with the flow of information and adaptive rules. One of the original aims of artificial neural networks (ANN) was to understand and shape the functional characteristics and computational properties of the brain when it performs cognitive processes such as sensorial perception, concept categorization, concept association and learning. However, today a great deal of effort is

focused on the development of neural networks for applications such as pattern recognition and classification, data compression and optimization.

Algorithms inspired by natural phenomena have taken center-stage in the development of new computer solutions. These algorithms have been found to be very efficient in solving complex computational problems such as optimizing objective functions, pattern recognition, control objectives, image processing and filter modeling. There exists a lot of literature on finding solutions for optimization problems. However, classical optimization algorithms such as Brute-Force Search, Breath- First Search, Uniform Cost Search and Depth First Search, do not provide suitable solutions to problems in pattern recognition, optimization of objective functions and image processing that require high-dimensional search space and complexity. In addition, techniques such as exhaustive search are not practical solutions to such problems (Flake, 1999).

There has been a large body of work in the area of swarm intelligence for optimization and solving different problems. Ant Colony Optimization (ACO) is one such algorithm that is based on the indirect communication between the ants by means of chemical pheromone trails, enabling them to find short paths between their nest and food sources. Other popular heuristic algorithms include the genetic algorithm (GA) inspired from the Darwinian evolutionary theory, Artificial Immune System (AIS) that simulates the immune system (Farmer et al., 1986), and Particle Swarm Optimization (PSO) based on simulation of the swarm behavior of a flock of birds. Zne and Chou (2005) proposed a hybrid search algorithm combining the advantages of GA and ACO that is able to explore the search space and exploit the best solutions. Dorigo et al. (1996) introduced the Ant System (AS), an analogy of ACO. The main characteristics of their model are positive feedback, cooperation by transferring information, and the use of a constructive greedy heuristic where agents compete to survive. Dong et al. (2007) modeled the social foraging behavior of Escherichia coli bacteria to solve optimization problems. They proposed a hybrid approach involving GA and bacterial foraging (BF) algorithms for function optimization problems. Rashedi et al. (2009) proposed the Gravitational Search Algorithm (GSA) based on the law of gravity and mass interactions. In their algorithm, the searcher agents are a collection of masses that interact with each other based on Newtonian gravity and the laws of motion. In another study, Formato (2008) introduced the Central Force Optimization (CFO), a new deterministic multi-dimensional search metaheuristic based on the metaphor of gravitational kinematics. Overall, the review of existing literature reveals that there is no one superior method for

solving optimization problems. This concept is proven by the "no free lunch theorems for optimization," which describes that no one algorithm is able to solve all the optimization algorithms but each algorithm can solve a special class of problems. Although many algorithms developed to solve optimization problems do achieve good performance, there are still some shortcomings. For example, the standard PSO algorithm often gets trapped in local optima when solving complex multimodal problems. GA has no absolute assurance of finding a global optimum and representation of the problem for GA is often difficult. Furthermore, there are various operations such as mutation and crossover in GA that require long response times in finding the solution for some problems. Evolutionary algorithms (EA) too suffer from the slow convergence problem.

The motivation for this work is the desire to improve measurement accuracy that would lead to development of better automatic methods for the purpose of detecting the heart signal disorders. Another challenge for biomedical signal processing is feature extraction to for understanding and identification the hidden information in a signal. These feature extraction methods can be designed to support the physician diagnosis by extracting information that is not easily available through simple visual assessment. Most of the current computerised methods aim to mimic the manual procedures that the physician follows for measuring the features in the signal. However, there are small measures in the heart signal that are not perceivable by the human eye and they are very useful for heart disorder detection and thus, need to be explored.

The premier audiences of the present book are graduate students and researchers who need to find a relationship between artificial intelligence and medical science and are interested to use informatics to improve the quality of the medical diagnosis approaches. Following is a brief description of each chapter.

The main objective of this research book is the automatic normalisation and classification of ECG signals of heart disorders. Two intelligent approaches based on Self-Organising Map (SOM) and Particle Swarm Optimization Neural Network (PSONN) are proposed to find the cutoff frequency for high frequency noise removal in ECG signals. 100 lead II ECG signals were obtained from the Physiobank database, composed of five types of ECG signals, including Normal, Supraventricular tachycardia, Bundle branch block, Anterior myocardial infarction (Anterior MI), and Inferior myocardial infarction (Inferior MI). For each ECG signal, the PSONN and SOM-identified cutoff frequency are applied to a Finite Impulse Response

(FIR) filter, and the resulting signal is evaluated against the original clean and conventionally filtered ECG signals. The results show that the proposed intelligent system successfully denoised the ECG signals more effectively than the conventional method.

With the help of a heart specialist, the features of the denoised signals were identified. The Kinetic Gas Molecule Optimization (KGMO) algorithm was developed to train the KGMO feedforward neural network (KGMONN) for classification of the ECG signals. The proposed KGMO was evaluated by using it to find the global minimum of 23 benchmark functions. Results compared against some of the benchmark and newest swarm based algorithms show that KGMO works significantly faster with better results. The statistical results for the KGMONN classification of ECG signals also show that the proposed classification algorithm outperformed other conventional methods. The performance evaluations indicate the effectiveness of the proposed intelligent algorithms and system for automatic classification of ECG signals for common heart disorders with the accuracy of 84%.

The book has 9 chapters, as follows:

Chapter 1 is an introduction on medical diagnosis and the way how artificial intelligence can help to improve disease diagnosis based on machine learning. Chapter 2 is the introduction that explains the function of the heart and an introduction to ECG signals. In addition, the importance of developing an intelligent system for heart disorder diagnosis and the background of the research is described. Chapters 3, 4, and 5 are introductions on Artificial Neural Networks, Swarm Intelligence and Feature extraction, respectively. Chapter 6 provides the literature review of the ECG signal waveform, and a brief description on neural network and swarm intelligence. Chapter 7 is the description of the methodology and details of two proposed algorithms for noise removal using SOM and PSONN. Chapter 8 concentrates on the proposed algorithm based on gas molecules behaviour and validation by testing on 23 benchmark functions. The results are compared with current algorithms to evaluate the performance of the proposed algorithm.

In Chapter 9, first feature extraction is explained and then the proposed KGMO is applied to a feedforward neural network to train it to classify the ECG signals of heart disorders. For evaluating the results, other training algorithms are also used to train the neural network for classification of ECG signals. The Conclusion gives the overall conclusion of the study and proposed future work to improve the performance of the system.

This is the place that I greatly say my thanks to many people from the first that I inspired from their knowledge and their idea to write this book. I am thankful for the comments of Prof. R. Logeswaran, Dr. Hasan S. Monadjemi for their technical comments for improving the exposition. I am also grateful to Dr. Zia Tadayon for his invaluable knowledge of heart. My last thanks but not the least belongs to the IGI Global team for their patience until finishing the book and thanks to the reviewers of my book for their comments.

Sara Moein
Mount Sinai School of Medicine, USA

REFERENCES

Adlam, D., & Hampton, R. (1997). *The ECG made easy*. Amsterdam: Elsevier.

Clancey, W. J., & Shortliffe, E. H. (1984). *Readings in medical artificial intelligence: The first decade*. Reading, MA: Addison Wesley.

Dong, H. K., Abraham, A., & Jae, H. C. (2007). A hybrid genetic algorithm and bacterial foraging approach for global optimization. *Information Sciences*, (177): 3918–3937.

Dorigo, M., Maniezzo, V., & Colorni, A. (1996). The ant system: Optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man, and Cybernetics*, *26*(1), 29–41. doi:10.1109/3477.484436 PMID:18263004

Farmer, J. D., Packard, N. H., & Perelson, A. S. (1986). The immune system, adaptation and machine learning. *Physica*, (D2), 187–204.

Flake, G. (1999). *The computational beauty of nature*. Cambridge, MA: MIT Press.

Formato, R. A. (2008). Central force optimization: A new nature inspired computational framework for multidimensional search and optimization. *Studies in Computational Intelligence*, (129), 221-238.

Rashedi, E., Nezamabadi-pour, H., & Saryazdi, S. (2009). GSA: A gravitational search algorithm. *Information Sciences*, *179*(13), 2232–2248. doi:10.1016/j.ins.2009.03.004

Zne, J. L., & Chou, Y. L. (2005). A hybrid search algorithm with heuristics for resource allocation problem. *Informatics and Computer Science*, (173), 155-167.

Acknowledgment

I am greatly thankful to many people that from the first inspired me from their knowledge and their ideas to write this book. I am also grateful to Dr. Zia Tadayon for his invaluable knowledge of the heart. This is a place to say thanks to Prof. R. Logeswaran, Dr. Hasan S. Monadjemi, and Prof. Hamid R. Arabnia for their technical comments for improving the exposition. My last thanks but not the least belongs to the IGI Global team for their patience until finishing the book and thanks to the reviewers of my book for their comments.

Chapter 1
Medical Diagnosis

ABSTRACT

This chapter briefly explains the medical diagnosis definition and the useful techniques that help to improve the performance of the existing medical diagnosis systems. The reasons for importance and difficulties of medical diagnosis and web-based medical diagnosis system components are explained and WISER as an example is provided.

1.1 INTRODUCTION

Whereas you need around five years to get your MSc or MA in other courses, you have to spend almost twice in a medical school to get your diploma. That simply suggests the complexity of the medical diagnosis. In each case, diverse symptoms generated by diverse causes should be considered and added to the patient background. Epidemics also should be considered and the genetic factors too, and then a diagnosis can be materialized. After a while, a physician would be more experienced, but just gradually and in one branch of diseases. However, there are still some problems. For many years, medical diagnosis was an art among human society. There are famous physicians as well as famous painters or composers throughout the history. Again, An artist is a person who can carry out something that others can not, and that is exactly what a professional physician does during a medical diagnosis procedure. Experts of medical science employ their educations, experiences, and talent, to diagnose a disease. A patient complaint is start of any diagnosis procedure and the expert learns more about the patient situation interactively during

DOI: 10.4018/978-1-5225-5580-3.ch001

an interview, as well as by measuring some metrics such as blood pressure or the body temperature.

Human is going to apply various methods in order to increase the quality of medical diagnosis. Especially computer-based approaches have important role in the area of medical services. There are two main categories for these computerized methods: first conventional techniques, such as database management systems (DBMSs), and second artificial intelligence (AI) techniques, such as knowledge-based systems (KBSs) or expert systems (ESs).

A part of medical diagnosis is related to diagnostic test that has seen an exponential increase in the accuracy and sensitivity in new decades. Diagnostic tests include observing external symptoms and using sophisticated laboratory tests and complex imaging methods that permit detailed non-invasive internal examinations. The improved methods for diagnostic tests cause providing accurate patient data to physician. A process of medical diagnosis is established when evidence to distinguish a probable cause of the patient's key symptoms from all other possible causes of the symptom is found. Various ranges of diseases such as cancer research, gastroenterology, heart diseases, use computer technology for medical decision support. Generically Decision Support Systems (DSS) are any type of application that support the decision-making process. A generic DSS receives a certain amount of data as input, processes it using a specific methodology and offers as a result some output that can help decision-makers (Caruana, R., 1996).

AI methods are applicable for medical data mining. Therefore, it provides a tremendous opportunity for data mining methods to assist the physician to collect the medical data of each patients and scientific knowledge. Physicians can use medical data mining in a variety of ways, by using interpret complex diagnostic tests, by combining information from multiple sources (sample movies, images, clinical data, proteomics, scientific knowledge), by providing support for differential diagnosis and providing patient-specific prognosis. Therefore, it is logical to claim that AI can be very helpful for solving the problems of real world. In other words, data mining can be considered as a branch of AI to be applied for information extraction from patients' data. Essentially, data mining gives information that would not be available otherwise. When the data collected involves individual people, there are many questions concerning privacy, legality and ethics.

Medical diagnosis is not a 100% reliable process. Either machine or a physician does mistake. However, it is true to think that all physicians do not make the same medical treatment. In spite of huge development of the

biomedical technology, the diagnostic accuracy is in many cases rather low. The reason is that medical instruments do not provide enough relevant information for reliable diagnosis. Recently developed technology for recording human's aura provide a completely new information about the biophysical and psychical state of the patient that could in some cases drastically improve the diagnostic process. However, the problem is the interpretation of the patient data. By using machine learning, we could alleviate that problem by means of automatically generating diagnostic rules from the records of patients with known diagnoses. Remember that not all the intelligent machines have the same performance and they do not have the same reliability. The more intelligent and experienced the machine is, the more reliable medical decision-making is gain.

Medical diagnosis is a very active field. It means that different research areas are supposed to be useful in this field in order to increase the performance of that. For instance, Database Management systems (DBMSs) are used for storing, retrieving and generally manipulating patient data, whereas expert systems (ESs) are mainly used for performing diagnoses based on patient data, since they can naturally represent the way experts reason. Diagnosis of bone diseases is greatly facilitated by the use of nuclear medicine methods, more specifically by the use of scintigraphic images (or scintigrams or scans), and a number of relevant expert systems have been developed (Altman, R. B., 1999).

The major problem in medical field is to diagnose disease. Human being always make mistake and because of their limitation, diagnosis would give the major issue of human expertise. One of the most important problems of medical diagnosis, in general, is the subjectivity of the specialist. It can be noted, in particular in pattern recognition activities, that the experience of the professional is closely related to the final diagnosis. This is due to the fact that the result does not depend on a systematized solution but on the interpretation of the patient's signal.

1.2 IMPORTANCE OF INTELLIGENT MEDICAL DIAGNOSIS

There are various problems in real world and some of them are more significant in comparison with others. Why medical diagnosis is one of these significant problems? The main reason is that medical diagnosis and treatment are related to human's life and body. Any mistake will face human life in danger. In addition, in case of some of the diseases, which related to heart, brain, or

different types of censers, medical diagnosis becomes more important. Any disease has its own unique symptoms. This matter makes physician's job to become more difficult and it increases the probability of wrong diagnosis.

Medical decision support system has an important rule in supporting the physicians in order to improve the quality of diagnosis. Any decision support system is composed of different parts such as patients data and clinical data and an intelligent engine to assign a disease to the data of patients. On the other hand, some patients disease is not so strong at start point and if the physician does not recognize the disease correctly at the suitable time then the symptoms becomes serious and the disease will develop toward sever problem. Even it can become an untreatable disease for all years of the patient's life. There is another point that without a proper medical diagnosis no proper treatment can be prescribed. All the mentioned points motivate us to pay more attention to do medical diagnosis supported by intelligent tools.

Because of the importance of using intelligent methods for increasing the accuracy of medical decision making systems, using artificial intelligence for medical diagnosis has become a major task. In 1984, Clancey and Shortliffe provided the following definition: 'medical artificial intelligence is primarily concerned with the construction of AI programs that perform diagnosis and make therapy recommendations. Unlike medical applications based on other programming methods, such as purely statistical and probabilistic methods, medical AI programs are based on symbolic models of disease entities and their relationship to patient factors and clinical manifestations. Actually, researchers studied a lot in applying artificial intelligence in medicine and they combined sophisticated representational and computing algorithms with the insights of expert physicians to design methods and tools for improving health care (Coiera, E. 2003).

The major task of medical science is to prevent and diagnose the diseases. Here our focus is the second task, which as mentioned before, is not a direct and simple task at all. In 2001, Brause highlighted that almost all the physicians are confronted during their formation by the task of learning to diagnose. Here, they have to solve the problem of deducing certain diseases or formulating a treatment based on more or less specified observations and knowledge.

Medical diagnosis has been one of the most difficult issues of human in before and now and there are reasonable justifications for this fact. Below some certain difficulties of medical diagnosis that have to be taken into account are listed:

- The basis for a valid diagnosis, a sufficient number of experienced cases, is reached only in the middle of a physician's career and is therefore not yet present at the end of the academic formation.
- This is especially true for rare or new diseases where also experienced physicians are in the same situation as new comers.
- Principally, humans do not resemble statistic computers but pattern recognition systems. Humans can recognize patterns or objects very easily but fail when probabilities have to be assigned to observations.
- The quality of diagnosis is totally depends on the physician talent as well as his/her experiences.
- Emotional problems and fatigue degrade the doctor's performance.
- The training procedure of doctors, in particular specialists, is a lengthily and expensive one. So even in developed countries we may feel the lack of medical doctors.
- Medical science is one of the most rapidly growing and changing fields of science. New results disqualify the older treats, new cures and new drugs are introduced day by day. Even unknown diseases turn up every now and then. Therefore, a physician should always try hard to keep his/herself up to date.

An intelligent medical diagnosis system can provide assistance for patient and medical experts. The intelligent system can serve to improve the quality of medical diagnosis, increases patient compliance and minimizes iatrogenic disease and medical errors. Medical diagnosis applications use AI in order to reduce the cost, time, human expertise and error in medical decision making. Intelligent methods are used for deriving diagnostic rules automatically from the descriptions of the patients treated in the past for which the final diagnoses were verified. After deriving the diagnostic rules, physicians use them for making the more reliable prescription (Wan & Siraj, 2006).

An artificial brain is used for storing and retrieving all symptoms of diseases that all called data; of course, the proficiency is needed to be mixed with content of that brain to gain to the best medical diagnosis. After preparing the dataset some various intelligent methods such as swarm intelligence or artificial neural network solve the diagnosis problem.

There are some advantages for any medical decision support. Therefore, many users could accept them. Some of these benefits are such as the low cost to deliver the knowledge for clinical diagnosis, availability vast amount of information at the same time and good performance for diagnosis. In some

systems, the supporting includes of managing the web based medical systems. Some instances are MEDLINE, EMBASE, Cochrane Library, Inspec, and ISI databases that are source of information for medical decision supports systems. A Web-Based medical diagnosis and prediction consists of four components as following (Wan & Siraj, 2006):

- Databases: The databases consist of patients database and patients-disease database. Patients database will be used to store patient's information such as name, addresses, and others particulars details. Patients-disease database stored all the information about patients and their illness. The information stored in the database includes types of diseases, the treatments and other details about the test and administering therapy. Patients information are separated in a different database to enhance the patients records storage, so that other departments could use the records when the patients are referred to them. This method could prevent other departments or unauthorized users from accessing the information about patients diseases and provide a centralized information access for the patients records.
- Prediction module: Prediction module and diagnosis module are two of the main features in Web-Based Medical Diagnosis and Prediction. Prediction module utilizes neural networks techniques to predict patients illness or conditions based on the previous similar cases. Data from the patients and patients-disease database will be used for training and testing. The weight from the training will be stored to predict a new data fed into the system.
- Diagnosis module: Diagnosis module consists of expert experience dealing with the same cases. Communications between doctors or specialist from other region helps doctor for diagnosing the patient and provides appropriate treatment. In telemedicine, Multimedia and Internet (or computer network) are two of the main tools that support the collaboration and distribution of information. Multimedia is a combination of media such as text, audio, visual and graphics can be used in medical application such as in image transmission (X-Ray images, pictures, etc.).
- User interface: User interface composed of HTML forms that provide facilities for user to use the diagnosis system. These forms are related to patients information, symptoms, etc.

An example of a web-based system is WISER (Wireless Information System for Emergency Responders). WISER is a system designed to assist First Responders in hazardous material incidents. Developed by the National Library of Medicine, WISER provides a wide range of information on hazardous substances, including substance identification support, physical characteristics, human health information, and containment and suppression guidance (Chellappa, M., 1995). Some of the main characteristics of WISER are as following:

- Mobile support, providing First Responders with critical information in the palm of their hand.
- Comprehensive decision support, including assistance in identification of an unknown substance and, once the substance is identified, providing guidance on immediate actions necessary to save lives and protect the environment.
- Rapid access to the most important information about a hazardous substance by an intelligent synopsis engine and display called "Key Info".
- Radiological support, including radioisotope substance data, tools, and reference materials.
- Intuitive, simple, and logical user interface developed by working with experienced first responders.

This book mainly concentrates on using artificial intelligence for medical diagnosis specifically about detection of heart disorders. In the next chapter, there is an explanation about heart and some of its common problems and detecting it by signals.

1.3 SUMMARY

Medical diagnosis has been one of the most difficult issues of human in before and now and there are reasonable justifications for this fact. It is not a 100% reliable process. Either machine or a physician does mistake. However, it is true to think that all physicians do not make the same medical treatment. Human is going to apply various methods in order to increase the quality of medical diagnosis. Especially computer-based approaches have important role in the area of medical services. An intelligent medical diagnosis system can provide assistance for patient and medical experts. Intelligent methods

are used for deriving diagnostic rules automatically from the descriptions of the patients treated in the past for which the final diagnoses were verified. After deriving the diagnostic rules, physicians use them for making the more reliable prescription.

REFERENCES

Alexopoulos, E., Dounias, G. D., & Vemmos, K. (1999). Medical Diagnosis of Stroke Using Inductive Machine Learning. *Machine Learning and Applications: Machine Learning in Medical Applications*, 20-23.

Amato, F., Lopez, A., Maria, E., Vanhara, P., Hampl, A., & Havel, J. (2013). Artificial Neural Network in Medical Diagnosis. *Journal of Applied Biomedicine*, *11*(2), 47–58. doi:10.2478/v10136-012-0031-x

Bonfa, I., Maioli, C., Sarti, F., Milandri, G. L., & Monte, P. R. D. (1993). *HERMES: An Expert System for the Prognosis of Hepatic Diseases*. Technical Report UBLCS-93-19. Universiti of Bologna.

Brause, R. W. (2001). Medical Analysis and Diagnosis by Neural Networks. J.W. Goethe-University.

Caruana, R., Baluja, S., & Mitchell, T. (1996). Advances in Neural Information Processing Systems: Vol. 8. *Using the Future to "Sort Out" the Present: Rankrop and Multitask Learning for Medical Risk Evaluation*. Cambridge, MA: The MIT Press.

Chellappa, M. (1995). Telemedic-Care. In *NCIT'95: 8'th National Conference Information Technology'95*. Gabungan Komputer Nasional Malaysia.

Coiera, E. (2003). *Guide to Medical Informatics, the Internet and Telemedicine*. Academic Press.

Detmer, W. M., & Shortliffe, E. H. (1997). Using the Internet to Improve Knowledge Diffusion in Medicine. *Communications of the Associations of Computing Machinery*, *40*(8), 101–108. doi:10.1145/257874.257897

Hoong, N. K. (1988). Medical Information Science - Framework and Potential. In *International Seminar and Exhibition Computerization for Development-the Research Challenge*. Universiti Pertanian Malaysia.

Jankowski, N. (1999). Approximation and Classification in Medicine with IncNet Neural Networks. In *Machine Learning and Applications: Machine Learning in Medical Applications*, 53-58.

Jones, M. T. (2005). *AI Application Programming* (2nd ed.). Hingham, MA: Charles River Media Inc.

Karabatak, M., & Cevdet-Ince, M. (2009). An expert system for detection of breast cancer based on association rules and neural network. *Expert Systems with Applications*, *36*(2), 3465–3469. doi:10.1016/j.eswa.2008.02.064

Kononenco, I. (2001). Machine Learning for Medical Diagnosis: History, State of the Art and perspective. *Artificial Intelligence in Medicine*, *23*(1), 89–109. doi:10.1016/S0933-3657(01)00077-X PMID:11470218

Kumar, A. V. S. (2012). Diagnosis of Heart Diseases Using Fuzzy Resolution Mechanism. *Journal of Artificial Intelligence*, *5*(1), 47–55. doi:10.3923/jai.2012.47.55

Machado, L. O. (1996). *Medical Applications of Artificial Neural Networks: Connectionist Model of Survival* (Ph.D dissertation). Stanford University.

Mahabala, H. N., Chandrasekhara, M. K., Baskar, S., Ramesh, S., & Somasundaram, M. S. (1992). ICHT: An Intelligent Referral System for Primary Child Health Care. *Proceedings SEARCC*, *92*(11).

Manickam, S., & Abidi, S. S. R. (1999). Experienced Based Medical Diagnostics System Over The World Wide Web (WWW). *Proceedings of The First National Conference on Artificial Intelligence Application In Industry*, 47 - 56.

Meng, Y. K. (1996). Interval-Based Reasoning in Medical Diagnosis. *Proceedings of National Conference on Research and Development in Computer Science and Its Applications (REDECS'96)*, 220 - 224.

Moein, S., Monadjemi, S. A., & Moallem, P. A. (2008). Novel fuzzy-neural based medical diagnosis system. *Int J. Biological and Life Sciences*, *4*(3), 146–150.

Murdoch, T. B., & Detsky, A. S. (2013). The inevitable application of big data to health care. *Journal of the American Medical Association*, *309*(13), 1351–1352. doi:10.1001/jama.2013.393 PMID:23549579

Neill, D. B. (2013). Using artificial intelligence to improve hospital inpatient care. *IEEE Intelligent Systems*, *28*(2), 92–95. doi:10.1109/MIS.2013.51

Patel, V. L., Shortliffe, E. H., Stefanelli, M., Szolovits, P., Berthold, M. R., Bellazzi, R., & Abu-Hanna, A. (2009). The coming of age of artificial intelligence in medicine. *Artificial Intelligence in Medicine*, *46*(1), 5–17. doi:10.1016/j.artmed.2008.07.017 PMID:18790621

Peng, L., Yang, B., & Jiang, J. (2009). A novel feature selection approach for biomedical data classification. *Journal of Biomedical Informatics*, *179*(1), 809–819. doi:10.1016/j.jbi.2009.07.008 PMID:19647098

Pomi, A., & Olivera, F. (2006). *BMC Medical Informatics and Decision Making, Context-sensitive auto associative memories as expert systems in medical diagnosis*. BioMed Central.

Shortliffe, E. H. (1987). Computer Programs to Support Clinical Decision Making. *Journal of the American Medical Association*, *258*(1), 61. doi:10.1001/jama.1987.03400010065029 PMID:3586293

Shortliffe, E. H. (1998). *The Evalution of Health-Care Records in the Era of the Internet*. Retrieved from http://smi-web.stanford.edu/pubs/abstracts_by_author/

Shortliffe, E. H., & Buchanan, B. G. (1975). A Model of Inexact Reasoning in Medicine. *Mathematical Biosciences*, *23*(3-4), 351–379. doi:10.1016/0025-5564(75)90047-4

Street, W. N., Mangasarian, O. L., & Wolberg, W. H. (1996). Individual and Collective Prognostic Prediction. *Thirteenth International Conference on Machine Learning*.

Thakur, A., Mishra, V., & Jain, S. (2011). Feed forward artificial neural network: Tool for early detection of ovarian cancer. *Scientia Pharmaceutica*, *79*(3), 493–505. doi:10.3797cipharm.1105-11 PMID:21886899

Theodorou, T., & Ketikidis, P. (1995). Neo-Dat An Expert System to Support the Designers of Clinical Trails. *5th Hellenic Conference on Informatics*.

Übeyli, E. D. (2007). Implementing automated diagnostic systems for breast cancer detection. *Expert Systems with Applications*, *33*(4), 1054–1062. doi:10.1016/j.eswa.2006.08.005

Wan, H. I. W., & Siraj, F. (2006). *Artificial intelligence in medical applications: an exploration*. Universiti Utara Malaysia.

Warner, D. (1997). *Malaysian Medical Matrix: Telemedicine in the age of the Multimedia Super Coridor*. Academic Press.

Yana, H., Jiangb, Y. T., Zhenge, J., Pengc, C., & Lid, Q. (2006). A Multilayer Perceptron-Based Medical Decision Support System for Heart Disease Diagnosis. *Expert Systems with Applications*, *2*(30), 272–281. doi:10.1016/j.eswa.2005.07.022

Zelic, I., Lavrac, N., Najdenov, P., & Rener-Primec, Z. (1999). Impact of machine Learning of the Diagnosis and Prognosis of First Cerebral Paroxysm. *Machine Learning and Applications: Machine Learning in Medical Applications,* 24-26.

Zrimec, T., & Kononenko, I. (2004). *Feasibility Analysis of Machine Learning in Medical Diagnosis from Aura Images*. University of Ljubljana.

Chapter 2
Introduction to Heart

ABSTRACT

This chapter provides an introduction to the heart and the importance of detecting heart problems based on heart signals. It explains details about electrocardiogram signal and 4 common heart disorders including supraventricular tachycardia, bundle branch block, anterior myocardial infarction (Anterior MI), and inferior myocardial infarction (Inferior MI).

2.1 OVERVIEW

Heart is an important organ in the body, for pumping the blood throughout the human body. The heart generates signals, which hide information in their structure. This signal information is very useful to physicians for heart disorder detection, but is not easily perceived by them. Therefore, it is necessary that signals be decoded to be useful for interpretation by specialists. The interpretation process is sometimes easy when it only involves visual inspection of the signal. However, there are signals whose complexity is high due to the heart disorders, which affect the decoded form of signals. Automated methods can assist specialists in accurate diagnosis and therefore, automated heart signal processing becomes a reliable tool for finding important information that is hidden in the signal (Adlam and Hampton, 1997).

An important issue in designing an analysis method for heart disorder diagnosis and avoiding the risk of overlooking useful information of the heart signals is to have the knowledge about the physiology of the human body. Early use of computers in the medicine area was limited to the automation,

DOI: 10.4018/978-1-5225-5580-3.ch002

but this goal has been changed over the years, because specialists must be responsible for the taken diagnosed disorders. Nowadays, the automated system goal is to help physicians for better decision making (Fitzgibbon et al. 2002).

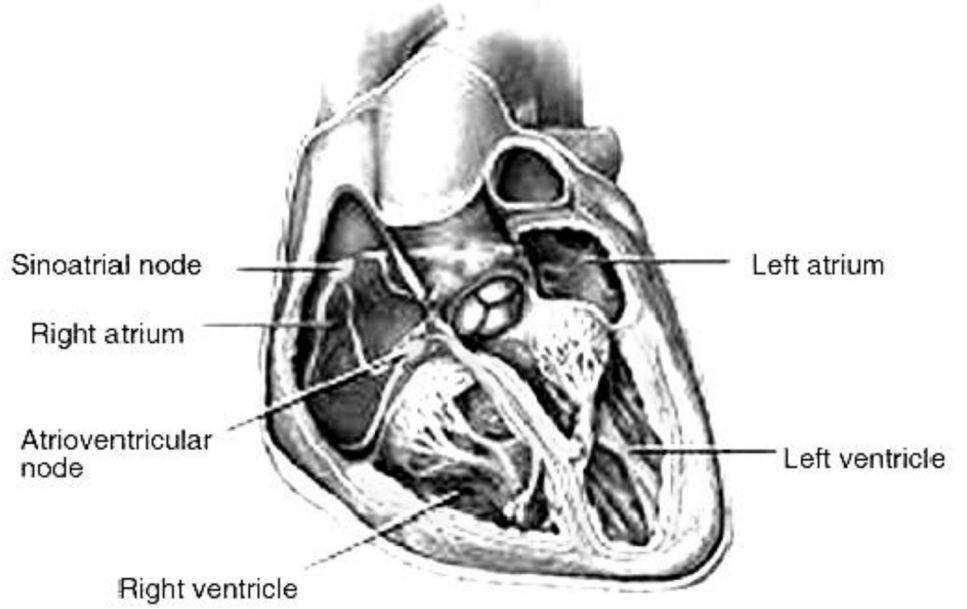
An emphasis of the book is noise removal as the heart signal is usually corrupted with noise from various sources, including machine malfunction, electrical noise from elsewhere in the body, respiration and muscle contractions (Behbahani, 2007). The noise consists of low-frequency and high-frequency components such as baseline wander and powerline interference, respectively (Karl et al., 2004). The recorded signal is distorted in a way that it could be difficult to perform any automatic diagnosis. The next sections present explanation on the function of the heart, the electrocardiogram, objectives, aims and contributions of this research work.

2.2 HEART FUNCTION

The heart is an organ in the human body for providing blood and oxygen. It is divided into 2 halves containing four chambers, as shown in Figure 1. As seen, left and right atria are upper chambers, while the left and right ventricles are the lower chambers. There are fibrous, non-conductive tissues for joining the atria to the ventricles to keep the ventricles electrically isolated from the atria. In addition, the heart contains veins called the superior and inferior vena cava for receiving the oxygen-poor blood into the right atrium. In order to pump the blood to the lung, the right atrium and the right ventricle cooperate together. The blood is forced into the right ventricle by the right atrium. In order to oxygenate the blood, the right ventricle then pumps it to the lungs. The oxygen-enriched blood received from the lung by the left atrium and the left ventricle circulates to the rest of the body (Adlam and Hampton, 1997).

Regular electrical impulses in the heart are spontaneously generated by the node called heart Sinoatrial (SA). The electrical impulses help the heart conduction system to initiate the contraction of the myocardium. In the heart function, there is a process called depolarisation that is because of the propagation of an electrical impulse to through the heart tissue. The depolarisation of the heart muscles causes generation of a strong ionic current (Silver, 2002). The generated current provides a voltage drop by flowing through the resistive body tissue. The electrical impulse flows to the atrial myocardium as a result of atrial depolarisation. The electrical impulse is

Figure 1. The heart conduction system (Adlam, D., and Hampton, R.,1997)

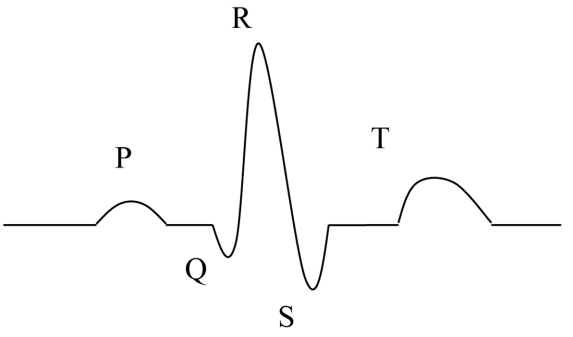


spread throughout the ventricular myocardium as the result of ventricular depolarisation.

2.3 ELECTROCARDIOGRAM (ECG)

Electrocardiogram (ECG) is a diagnosis tool that reports the electrical operation of the heart, recorded by skin electrodes at specific locations on the body. An effective way to detect heart disorders is through the analysis of the ECG signals, generated by polarisation and depolarisation of the heart that occurs when pumping blood throughout the human body (Wiley et al., 2005). The ECG signal waveform consists of the *P* wave, *QRS* wave and *T* wave, as shown in Figure 2. *P* wave is the result of atrial depolarisation, the ventricular depolarisation is presented in the *QRS* complex and the ventricular repolarisation is represented in T wave. In the overall ECG interpretation for computerised programmes and the human electrocardiographer, it is vital to access accurate measurements of ECG intervals and axes (*P/ QRS/ T*). The morphology and heart rate reflects the cardiac health of the human heart beat

Figure 2. A typical waveform of ECG signal



(Tomas & Neil, 2002). It is a non-invasive technique to help the physician identify heart diseases.

Physicians use the ECG to detect cardiac arrhythmia by changes in the morphology of the heart beat pattern. The amplitude and duration of the *P/QRS/T* wave contain useful information about the nature of the heart disease. The polarisation and depolarisation of Na^+ and K^- ions in the blood cause the generation of heart electrical wave. The following information of the human heart is provided by the ECG signal (Silver, 2002):

- The chamber size and position of heart,
- Impulse propagation and origin,
- The rhythm of heart and disturbance in conduction,
- The location of heart infarction,
- Electrolyte concentrations changes,
- Effects of drug on heart.

ECG however does not afford data on cardiac contraction or pumping function.

2.3.1 ECG Standard 12 Leads

A pair of electrodes (positive voltage and negative voltage) in contact with the body at designated anatomical locations is called a lead, which is then connected to an ECG record. There are 12 standard ECG leads including 3 bipolar leads, 3 augmented unipolar leads and 6 chest leads (Tomas & Neil, 2002).

Each type of lead is used for a designated purpose. Unipolar leads record the electrical potential at a particular point by means of a single exploring electrode, while the bipolar leads can record the potential difference between two points (positive voltage and negative voltage poles). The chest leads mainly detect potential vectors directed towards the back. These vectors are hardly detectable in the frontal plane (Gacek and Pedrycz, 2012). Table 1 provides the information of 12 ECG leads.

Leads I, II and III are called Einthoven leads for recording potentials between the left and right arms, between the right arm and left leg, and between the left arm and left leg, respectively.

Lead I is constructed by comparing the left arm (as positive) to the right arm's electrode (as negative). Lead I gives a very good view of what is going on from left to right in the heart, but a poor view of events moving up or down. Figure 3a shows the connections in lead I. The zero point is in the centre of the lead. Any current flowing to the left (towards the left arm's electrode) produces a positive deflection on the ECG, while any current flowing to the right produces a negative deflection. In essence, we are viewing the heart as if from the zero point (indicated by the sign *Zp*) (Tomas & Neil, 2002).

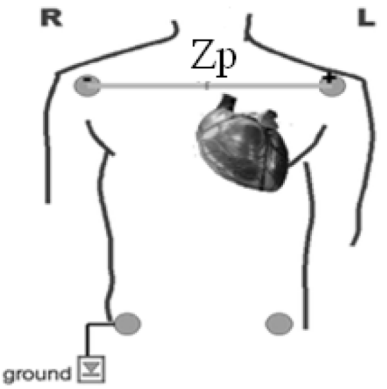
Lead II (Figure 3b) connects the left leg as positive to the right arm's negative. Lead III connects the left leg as positive to the negative left arm (Figure 3c). Each of these provides a view of the heart essentially from the zero point, as with lead I.

The unipolar augmented limb leads are called Goldberger leads placed in the frontal plane. The positive terminal can be placed on the right arm (aVR), left arm (aVL) or left leg (aVF). One lead connected to the positive terminal acts as the different electrode, while the other two limbs are connected to

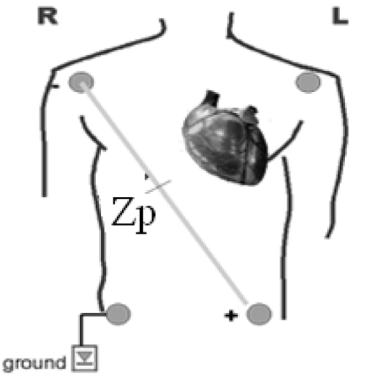
Table 1. ECG 12 leads

Standard Leads	Limb Leads	Chest Leads
Bipolar leads	Unipolar leads	Unipolar leads
Lead I Lead II Lead III	aVR aVL aVF	V1 V2 V3 V4 V5 V6

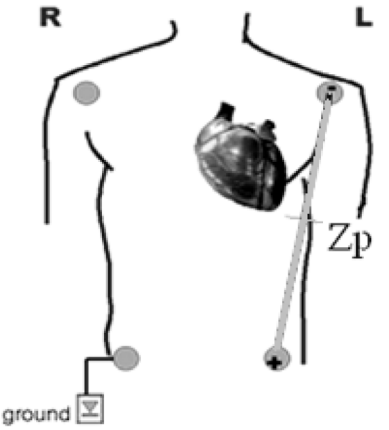
Figure 3. Position of standard bipolar leads



(a) Lead I



(b) Lead II



(c) Lead III

the negative terminal serve as the indifferent (reference) electrode (Tomas & Neil, 2002). Figure 4 shows the unipolar leads connected to the body.

The six leads (V1–V6) are Wilson leads that are unipolar chest leads are in an almost horizontal plane as shown in Figure 5. The indifferent electrode is obtained by connecting the three standard limb leads (Adlam and Hampton, 1997).

Figure 4. Position of unipolar limb leads

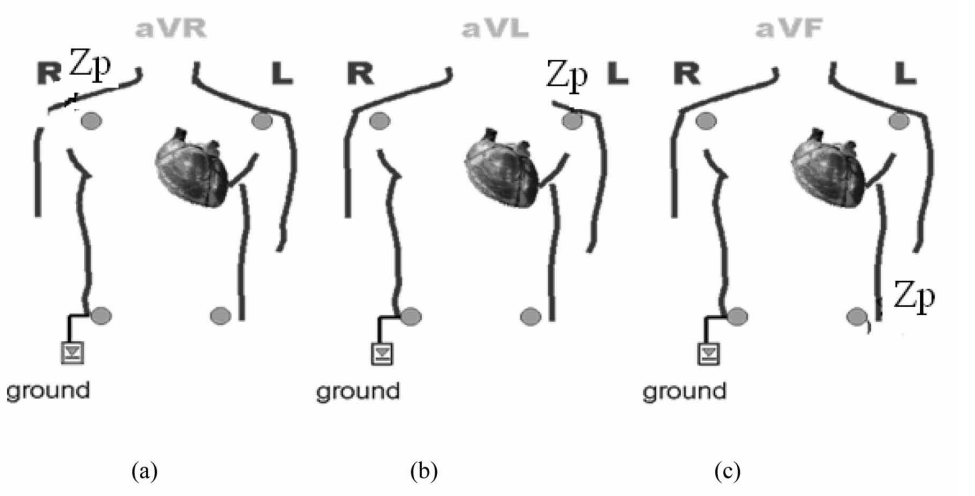
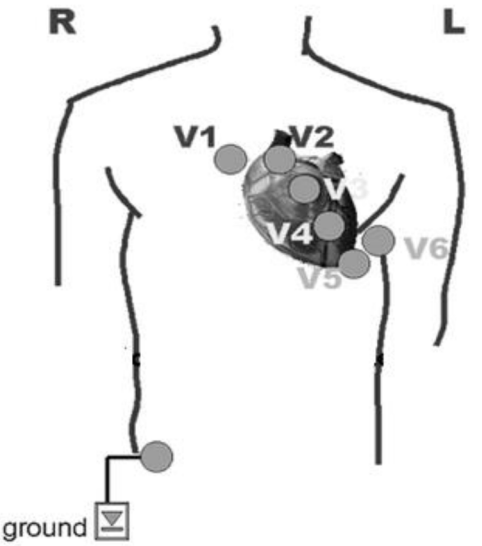


Figure 5. Unipolar chest leads



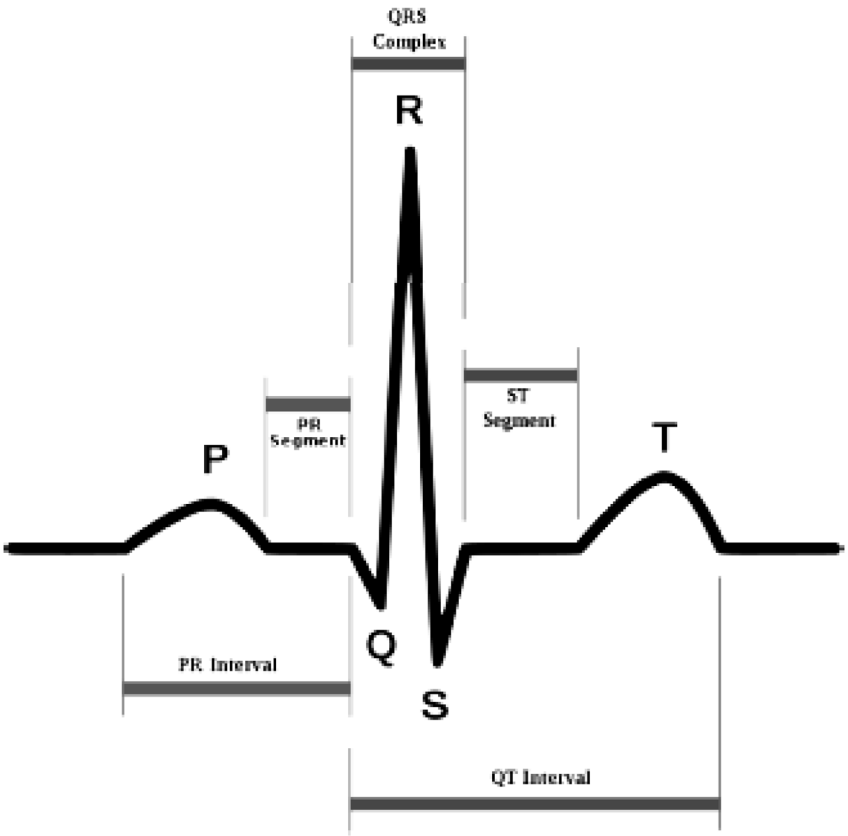
The chest leads (different electrode) can be used for recording. For forming an indifferent electrode with high resistance, three limb leads are connected. The QRS vector is usually directed towards and downwards the left back region, therefore the QRS vectors recorded by leads V1–V3 are usually negative, while those detected by V5 and V6 are positive. Since the chest

electrode in lead V1 and V2 is nearer to the base of the heart, which is the direction of electronegativity during most of the ventricular depolarisation process, the QRS is negative voltage. Because the chest electrode in V4, V5 and V6 leads are nearer the heart apex, the QRS is a positive voltage, which is the direction of electropositivity during most of the depolarisation.

2.3.2 The P/QRS/T Waveform

Figure 6 shows an example of ECG waveform. The ECG waveform can be broken down into three important parts represented by P, QRS and T, each denoting a peak. Each part of the ECG waveform represents a vital process in the heart (Gacek and Pedrycz, 2012). In case of an afflicting heart disorder,

Figure 6. Schematic representation of normal ECG waveform with intervals



the area that is not functioning normally would distort the wave. Therefore, the disease can be detected by inspection of the ECG wave form.

The heart intervals and segments are described as below (Tomas & Neil, 2002):

- **RR Interval:** RR is the inverse of heart rate and is the time duration between one R wave and the next R wave. The RR interval is the inverse of the heart rate. Normally its duration is 0.6s to 1.2s. A normal heart beating is between 50 and 100 beat per minute (bpm).
- **P Wave:** The P wave is as the result of directing the electrical impulse between AV node and SA node and spreads from the right atrium to the left atrium during normal atrial depolarisation. The normal duration of a P wave is 80 ms.
- **PR Interval:** It is the interval from beginning of T wave to the beginning of QRS complex. It appears when the electrical impulse propagates from the sinus node through the AV node and enters the ventricles. This travel is reflected in the PR interval. The normal duration is 120ms to 200ms.
- **PR Segment:** The P wave and the QRS complex are connected through the PR segment. The PR interval is a good estimation of atrioventricular (AV) node operation. The normal duration is 50ms to 120ms.
- **QRS Complex:** The rapid depolarisation of the right and left ventricles is reflected in the QRS complex. The mass of ventricular muscles is much larger than the atria muscles; therefore the amplitude of the QRS complex is larger than the P wave. The normal duration is 80ms to 120ms.
- **J Point:** The point of beginning the ST segment at which the QRS complex finishes. The degree of elevation or depression of the ST segment is presented by the J point.
- **ST Segment:** The QRS complex and the T wave are connected together by the ST segment, which is the period of depolarisation of the ventricles. The normal duration is 80ms to 120ms.
- **T Wave:** The T wave shows the repolarisation of the ventricles. It is also called recovery of ventricles. The beginning of the QRS complex to the apex of the T wave composes an interval that reflects the refractory duration. The normal duration is 160ms.
- **ST Interval:** The ST interval is the duration from the J point to the end of the T wave. The normal duration is 360ms.

- **QT Interval:** QT interval is the duration between the beginning of the QRS complex to the end of the T wave represents the QT interval. When the QT interval prolongs, there will be a danger for sudden death. The normal duration is 300ms to 400ms.

By inspection of the ECG waveform, the heart specialist can detect any disorder. The next section describes of the disorders explored in this research work.

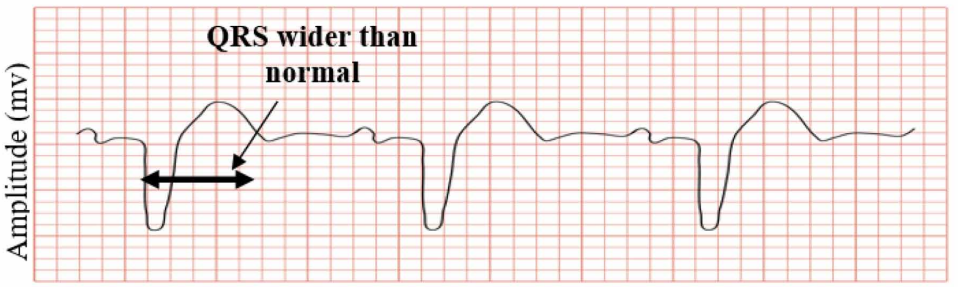
2.4 COMMON HEART DISORDERS

Any disturbance in the rate, regularity, and site of origin or conduction of the cardiac electric impulse causes heart arrhythmia (Adlam and Hampton, 1997). This section illustrates some of the common heart disorders.

2.4.1 Bundle Branch Block

Bundle branch block (BBB) makes it difficult for the heart to effectively pump the blood. In most people, there is no cause or symptom for BBB, but for those who have some symptoms it, can include fainting (syncope), feeling as if about go to faint (presyncope), or having slow heart beat. BBB happens when there is a cease in the conduction of the impulse from the AV node to the whole conduction system of the heart. The myocardial infarction or cardiac surgery also can occur as the result of this block (Karl, et al. 2004).

Figure 7. ECG Bundle branch block



Diagnosis of a BBB is possible by ECG as the duration of the QRS complex

exceeds 120 ms. Figure 7 shows an example ECG of the BBB where the QRS is wider than the normal duration (120 ms).

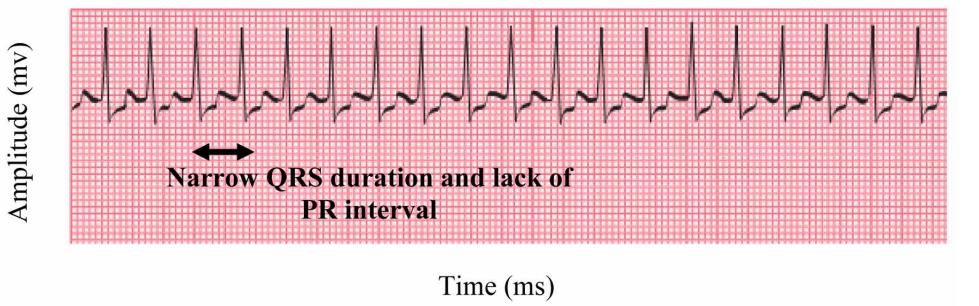
There are two types of BBB beats. The left BBB beat (LBBB) and right BBB beat (RBBB). In the left bundle branch block the electrical impulse is not able to propagate from the AV node from depolarising the left ventricular myocardium in the normal way, LBBB occurs. In RBBB, the electrical impulse is prevented from the AV node to the conduction network to depolarise the right ventricular myocardium (Wiley et al., 2005).

2.4.2 Supraventricular Tachycardia

Supraventricular means "above the ventricles" originating from the atria. Supraventricular tachycardia (SVT) is one type of heart rhythm disorder in which the heart beats between 140 to 220 times a minute. Therefore, the RR interval is narrower than normal. Commonly with most of the SVTs, the QRS duration is narrowed in the ECG (Caldwell, 2007). Figure 8 shows that the SVT that has a narrow QRS and narrow RR interval. Sometimes, it is referred to as atrial tachycardias. Some of the symptoms of SVT include palpitations, anxiety, light-headedness, chest pain, pounding in the neck and chest, and dyspnea. SVT is not usually related to the heart muscles disease, and is mostly associated with excessive intake of caffeine or alcohol.

Heart specialists classify the SVT based on the path that the electrical signal takes from the atria. Below are some types of SVT (Caldwell, 2007):

Figure 8. ECG supraventricular tachycardia



- AV nodal reentrant tachycardia (AVNRT) is a type of SVT that extra fiber in and around the AV node generates circular travelling of electrical impulse.
- Bypass tract or accessory pathway is an electrical conduction via extra fiber between the atria and ventricles that produces another type of SVT which is called AV reentrant tachycardia, or AVRT. In AVRT, the electrical impulse travels down the AV node to the ventricle and back to the atrium via these extra fiber, producing the SVT.
- Atrial tachycardias are another type of atrial disorder that occurs when the localized regions in the atria develop the ability to fire rapidly on their own.

2.4.3 Inferior and Anterior Myocardial Infarction

Another name for myocardial infarction (MI) is heart attack. MI occurs when a coronary artery has become blocked and it causes preventing the blood flowing to an area of myocardium. The severity of the infarct may have different effects on the layers of myocardium. Fibrous tissue is formed within the necrotic area. The location and the amount of affected tissue determine the amount of the effect on each patient. The problem occurs when the fibrous / necrotic area is unable to contract or conduct electrical impulses. This causes reduction in the efficiency of the chambers (Caldwell, 2007). During an infarct, there is normally ST elevation and T wave inversion. There may also be the formation of Q waves.

Mainly, there are various types of MI as given below:

1. **Anterior Wall:** It occurs due to an occlusion of the left anterior descending artery. The disorder is called Anterior MI, which affects the front wall of the left ventricle. In some cases the papillary muscles and intraventricular septum may be affected (anteroseptal). This type of infarction is very dangerous for a patient as it reduces the cardiac output dramatically. The ECG changes associated with anterior MI, including pathologic Q waves and repolarisation abnormalities, are best seen in at least two subsequent or contiguous leads from V1 to V5. New-onset poor R wave progression may be seen in patients with old anterior MI. In these patients, previous ECG tracings showing normal R wave progression suggest that the new-onset poor R wave progression is due to the anterior MI.

2. **Lateral Wall:** The occlusion of the lateral branch of the circumflex artery will cause the Lateral MI. Sometimes the patient has the occlusion in the left anterior descending artery that will result in an Anterolateral MI.
3. **Inferior Wall:** The occlusion of the right coronary artery will cause the infarction in the left ventricular, which is called Inferior MI. The ECG changes associated with inferior MI, including pathologic Q waves and repolarisation abnormalities, are best seen in the inferior leads: II, III, and aVF.
4. **Posterior Wall:** when the posterior branch of the right coronary or left circumflex artery is occluded. In this disorder, the back wall of the left ventricle is affected.

Among the 4 types of MI, Anterior and Inferior MI are detectable by lead II, which are shown in Figure 9 and Figure 10, respectively. Figure 9 shows that long QT and inverted T are effects of inferior MI on ECG. Figure 10 shows that lack of Q wave and inverted T on the ECG because of anterior MI. Since typically the ECG signals are collected from lead II, and based on feedback from heart specialists, only those 2 types of MI are used as subjects of classification in this book.

Figure 9. Inferior MI

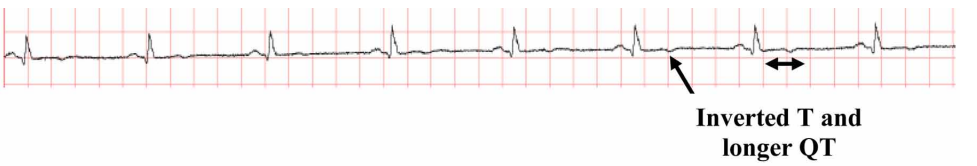
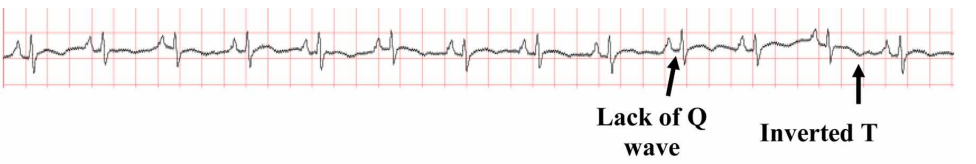


Figure 10. Anterior MI



2.5 MACHINE LEARNING FOR ELECTROCARDIOGRAM ANALYSIS

Detection of heart disorders using machine learning is one of the well-known topics among the researchers, which has some difficulties. ECG signals are corrupted with noise and artifacts that makes difficulties in analysis of signals (Alfaouri and Daqrouq 2008). Therefore, machine learning algorithms need to fulfill many requirements to detect noise and then diagnose the disorder. Recently many studies are conducted in the field of detecting heart disorders using machine learning. Several methods have been used for classification of arrhythmia such as Genetic algorithms, Fuzzy Logic, Self-Organizing Map, Bayesian, Hidden Markov Models and SVMs (Gupta & Chatur, 2012). Sao et. al studied the patterns of ECG signals for diagnosis and showed that effective diagnosis need computerized methods for detecting cardiac disease (Sao et al., 2015). Neural networks are used for classification after pre-processing, detection and feature extraction from the ECG signal. In the study by Gao et al. (2004) one diagnosis system is developed using artificial neural network to classify cardiac arrhythmias. The classifier is based on a Bayesian framework, which is formulated with the application of back propagation algorithm. The prediction accuracy in the developed system is about 90%. In a different study, Vishwa et al., (2011) proposed an automated method using ANN for classification of cardiac arrhythmia using multi-channel ECG recordings. The MIT-BIH arrhythmia database consisting of 48 recordings from 1975 and 1979 is used. For validation of this method, a database of Normal Sinus Rhythm (NSR) is used. One study (Sao et al., 2015) used ANN in ECG for diagnosing heart diseases in the most efficient manner. The researchers developed an ANN classifier for the same, wherein the parameters such as Spectral Entropy, Largest Lyapunov exponent and Poincare plot geometry are assessed along with the utilization of back propagation algorithm.

Anuradha and Reddy (2008) used neural networks for optimizing the classification of ECG signals. They developed an ANN-based classifier for cardiac arrhythmia. The classifier uses a combination of wavelets Back Propagation Algorithms (BPA). The accuracy of the proposed method is 90.56% is achieved and the classifier efficiently detects the abnormal activities. There was lack of datasets to validate the proposed method. In the research Jadhav et al., (2012) a new system based on ANN is developed for diagnosing heart diseases. For this purpose, the researchers used 12 lead ECG signal recordings data from UCI machine-learning repository datasets. This

algorithm is based on back propagation algorithm and momentum learning rule for classification. The evaluate the performance of the developed method Mean Squared Error (MSE), Receiver Operating Characteristics (ROC) and Area Under Curve (AUC) are used. Different datasets are used for validating the proposed method. The classification accuracy is obtained to be 86.67% and the sensitivity is 93.75%.

The main focus in this book is to develop an intelligent system to classify ECG signals for 4 common heart disorders, which are supraventricular tachycardia, bundle branch block, anterior myocardial infarction (Anterior MI), and inferior myocardial infarction (Inferior MI) as well as the normal healthy class. Since the ECG signals are usually corrupted with noise, therefore noise removal is necessary before any analysis can be carried on the signals. It is an important phase of the system. The accuracy of the classification and the speed of convergence for the proposed method are two major factors that should be optimised. Following chapters provide explanation about Artificial Neural Network (ANN), which is one of the main machine learning tools for classification and intelligent medical diagnosis. Also some of the optimization algorithms are introduced that are used for this purpose. The next chapter provides an introduction to artificial neural network.

2.6 SUMMARY

The function of the human body is based on signals of electrical, chemical or acoustic origin. Such signals provide information which may not be immediately perceived but which is hidden in the structure of the signal. The ECG is characterized by a series of waves whose morphology and timing provide information used for diagnosing diseases reflected by disturbances of the electrical activity of the heart. Leads I, II and III are called Einthoven leads for recording potentials between the left and right arms, between the right arm and left leg, and between the left arm and left leg, respectively.

REFERENCES

Adlam, D., & Hampton, R. (1997). *The ECG Made Easy*. Kindle Edition.

Al-kazemi, B., & Mohan, C. K. (2002). Training feedforward neural network using multi-phase particle swarm optimization. *Proc 9th Intl Conf Neural Infor Process*, (2), 2615-2619. 10.1109/ICONIP.2002.1201969

Arkin, R. (1998). *Behavior-Based Robotics*. Cambridge, MA: MIT Press.

Bai, Q. (2010). Analysis of particle swarm optimization algorithm. *Computer and Information Science*, *3*(1), 180–183. doi:10.5539/cis.v3n1p180

Behbahani, S. (2007). Investigation of adaptive filtering for noise cancellation in ECG signals. *2nd Intl Multi-Symp Comp Comput Sci*, 144–149.

Bekir, K. M., Osman, T., & Musa, A. (2003). A fuzzy clustering neural network architecture for multifunction upper-limb prosthesis. *IEEE Transactions on Biomedical Engineering*, *50*(11), 1255–1261. doi:10.1109/TBME.2003.818469 PMID:14619995

Beni, G., & Wang, J. (1989). Swarm intelligence in cellular robotics systems. *Proc. NATO Advanced Workshop on Robots and Biological System*, 102.

Bergh, F. V. D. (2001). *An analysis of particle swarm optimisers* (Thesis). University of Pretoria, Pretoria, South Africa.

Blondin, J. (2009). *Particle Swarm Optimization: A Tutorial*. Amsterdam University.

Bonabeau, E., Dorigo, M., & Theraulaz, G. (1999). *Swarm Intelligence: From Natural to Artificial Systems*. Oxford University Press.

Buczak, A., & Uhrig, R. (1996). Hybrid fuzzy-genetic technique for multisensory fusion. *Information Sciences*, *93*(3-4), 265–281. doi:10.1016/0020-0255(96)00078-3

Caldwell, B. E. Jr. (2007). *Prevent and Reverse Heart Disease*. Penguin Group.

Castro, B., Kogan, D., & Geva, A. B. (2000). ECG feature extraction using optimal mother wavelet. *The 21st IEEE Convention of the Electrical and Electronic Engineers*, 346-350. 10.1109/EEEI.2000.924422

Ceylan, R., Yuksel, O., & Bekir, K. A. (2009). Novel approach for classification of ECG arrhythmias: Type-2 fuzzy clustering neural network. *Expert Systems with Applications*, *36*(3), 6721–6726. doi:10.1016/j.eswa.2008.08.028

Chazal, P., & Celler, B. G. (1998). Selecting a neural network structure for ECG diagnosis. *20th Ann Intl Conf IEEE – EMBS*, 1422-1425.

Chun-an, L. (2008). New dynamic constrained optimization PSO algorithm. *Proc. Fourth International Conference on Natural Computation*, 650-653.

Cong, L., Sha, Y., & Jiao, L. (2006). Numerical optimization using organisational particle swarm algorithm. Simulated Evolution and Training. *Lecture Notes in Computer Science*, (4247), 150–157. doi:10.1007/11903697_20

Corne, D., Dorigo, M., & Glover, F. (Eds.). (1999). *New Ideas in Optimization*. McGraw Hill.

Devi, S., Jadhav, D. G., & Pattnaik, S. S. (2011). PSO based memetic algorithm for unimodal and multimodal function optimization. Swarm, Evolutionary, And Memetic Computing. In *Proceedings of International Conference on Swarm, Evolutionary, and Memetic Computing*. New York: Springer. 10.1007/978-3-642-27172-4_16

Dong, H. K., Abraham, A., & Jae, H. C. (2007). A hybrid genetic algorithm and bacterial foraging approach for global optimization. *Information Sciences*, (177), 3918–3937.

Dorigo, M., Maniezzo, V., & Colorni, A. (1996). The ant system: Optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man, and Cybernetics*, *26*(1), 29–41. doi:10.1109/3477.484436 PMID:18263004

Eberhart, R., & Shi, Y. (2001). Particle swarm optimization: developments, application and resources. *Proceedings of IEEE Congress of Evolutionary Computation*, 81-86. 10.1109/CEC.2001.934374

Eberhart, R., Shi, Y., & Kennedy, J. (2001). *Swarm Intelligence*. Academic Press.

Hopwood, S. J., & Jeans, J. (2009). *An Introduction to the Kinetic Theory of Gases*. Cambridge University Press.

Hosseini, H. G., Reynolds, K. J., & Powers, D. (2001). A multi-stage neural network classifier for ECG events. *Proceedings 23rd Intl Conf IEEE EMBS*, (2), 1672-1675.

Hu, G. M., Palreddy, S., & Tompkins, W. (1997). Patient adaptable ECG beat classifier using a mixture of experts approach. *IEEE Transactions on Biomedical Engineering*, (44): 891–900. PMID:9282481

Idoumghar, L., Melkemi, M., Schott, R., & Aouad, M. I. (2011). *Hybrid PSO-SA type algorithms for multimodal function optimization and reducing energy consumption in embedded systems. Applied Computational Intelligence and Soft Computing*. Hindawi.

Inan, O. T., Giovangrandi, L., & Kovacs, G. T. A. (2006). Robust neural network based classification of premature ventricular contractions using wavelet transform and timing interval features. *IEEE Transactions on Biomedical Engineering*, *53*(12), 2507–2515. doi:10.1109/TBME.2006.880879 PMID:17153208

Ismail, W., & Shamsuddin, S. M. (2008). *Differential evolution for neural networks training enhancement* (M.Sc. Thesis). University Technology of Malaysia.

Jesmin, K., Sharif, B., and Gregory, M. (2011). Empirical mode decomposition of the ECG signal for noise removal. *Optical Pattern Recognition*, (8055), 504-512.

Karl, G. R., Isis, A. W., Roberto, L., & Hakan, N. (2004). Fetal ECG waveform analysis. Best Practice & Research. *Clinics in Obstetrics and Gynaecology*, *3*(18), 485–514.

Kennedy, J., & Eberhart, R. (1995). A new optimiser using particle swarm theory. *Proceedings of the 6th International Symposium on Micro Machine and Human Science*, 39-43.

Khan, J., Wei, J. S., Ringnér, M., Saal, L. H., Ladanyi, M., Westermann, F., ... Meltzer, P. S. (2001). Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nature Medicine*, *7*(6), 673–679. doi:10.1038/89044 PMID:11385503

Kim, G. H., Yoon, J. E., An, S. H., Cho, H. H., & Kang, K. I. (2004). Neural network model incorporating: A genetic algorithm in estimating construction cost. *Building and Environment*, *39*(11), 1333–1340. doi:10.1016/j.buildenv.2004.03.009

Kirkpatrik, S., Gellat, C. D., & Vecchi, M. P. (1983). *Optimization by simulated annealing*. Academic Press.

Kohavi, R., & Provost, F. (1998). Glossary of Terms. *Machine Learning*, *30*(2/3), 271–274. doi:10.1023/A:1017181826899

Konstantinos, N. P., Dimitrios, H., & Jimmy, K. M. L. (2006). *ECG biometric recognition without fiducial detection*. Biometric Consortium Conference. IEEE Biometrics Symposium, Baltimore, MD.

Leonard, B. L. (2004). *The Kinetic Theory of Gases*. Dover Phoenix Editions.

Lian, Y., & Hoo, P. C. (2006). Digital elliptic filter application for noise reduction in ECG signal. *WSEAS Trans Electron*, *3*(1), 65–70.

Lian, Z., Zhu, F., Guan, Z., & Shao, X. (2008). The analysis of particle swarm optimization algorithm's convergence. *Proceedings of the 7th World Congress on Intelligent Control and Automation*, 623-628.

Mehmet, E. (2004). ECG beat classification using neuro-fuzzy network. *Patt Recog Lett*, *25*(15), 1715–1722. doi:10.1016/j.patrec.2004.06.014

Minami, K., Nakajima, H., & Toyoshima, T. (1999). Real-time discrimination of ventricular tachyarrhythmia with Fourier transform neural network. *IEEE Transactions on Biomedical Engineering*, *46*(2), 179–185. doi:10.1109/10.740880 PMID:9932339

Moein, S., Monadjemi, S. A., & Moallem, P. A. (2008). Novel fuzzy-neural based medical diagnosis system. *Int J. Biological and Life Sciences*, *4*(3), 146–150.

Murali, N. V. (2004). *Design of FIR filters with arbitrary responses. Seminar report*. Bombay: Electrical System Group.

Naghsh-Nilchi, A. R., & Kadkhodamohammadi, A. R. (2008). *Cardiac arrhythmias classification method based on MUSIC morphological descriptors, and neural network. EURASIP J Adv Sig Process*.

Orfanidis, S. J. (1996). *Introduction to Signal Processing*. Upper Saddle River, NJ: Prentice Hall.

Osowski, S., & Linh, T. H. (2001). ECG beat recognition using fuzzy hybrid neural network. *IEEE Transactions on Biomedical Engineering*, *48*(11), 1265–1271. doi:10.1109/10.959322 PMID:11686625

Owis, M. I., Youseff, A. B. M., & Kadah, Y. M. (2002). Characterisation of electrocardiogram signals based on blind source separation. *Medical & Biological Engineering & Computing*, *40*(5), 557–564. doi:10.1007/BF02345455 PMID:12452417

Ozbay, Y., Ceylan, R., & Karlik, B. (2006). A fuzzy clustering neural network architecture for classification of ECG arrhythmias. *Computers in Biology and Medicine*, *36*(4), 376–388. doi:10.1016/j.compbiomed.2005.01.006 PMID:15878480

Pareja-Tobes, P., Pelta, D., Sancho-Royo, A., & Verdegay, J. L. (2008). Search spaces representation in optimization problems. *Expert Systems with Applications*, *34*(4), 2891–2895. doi:10.1016/j.eswa.2007.05.046

Silver, M. (2002). *Success With Heart Failure: Help and Hope for Those with Congestive Heart Failure* (3rd ed.). Da Capa Press.

Sivanandam, S. N., & Deepa, S. N. (2007). *Introduction to genetic algorithms*. Springer.

Sornmo, L., & Laguna, P. (2005). *Bioelectrical Signal Processing in Cardiac and Neurological Applications*. Academic Press.

Sotos, J. M., Melendez, C. S., Salort, C. V., Abad, R. C., and Ibanez, J. J. R. (2007). A learning based Widrow-Hoff Delta algorithm for noise reduction in biomedical signals, Bio-inspired Modelling of Cognitive Tasks. *Lecturer Notes in Computer Science*, (4527), 377-386.

Sotos, J. M., Sanchez, C., Mateo, J., Alcaraz, R., Vaya, C., & Rieta, J. J. (2007). Neural networks based approach to remove baseline drift in biomedical signals. *11th Mediterr Conf Med Biomed Eng Comput 2007. IFMBE Proceedings*, *16*(2), 90–93.

Tomas, B. G., & Neil, E. H. (2002). *Introduction to 12-Lead ECG: The Art of Interpretation*. Jones and Bartlett Publishers.

Chapter 3
Definition of Artificial Neural Network

ABSTRACT

This chapter is an explanation of artificial neural network (ANN), which is one of the machine learning tools applied for medical purposes. The biological and mathematical definition of neural network is provided and the activation functions effective for processing are listed. Some figures are collected for better understanding.

3.1 BIOLOGICAL NEURAL NETWORK

In the nervous system of the living creatures, there are fluid-filled sacs which bound by a lipid bilayer for separating the intracellular contents from the extracellular space and they are called Neurons, or brain cells. Inside the body, Neurons are responsible to maintain a negative internal voltage, which is related to the extracellular space; ion channels and pumps maintain this potential difference. In most neurons of the central nervous system, spike is responsible to send the signals of neural activity, or rapid intracellular depolarization followed by repolarization; in order to adjust the neurons, it is necessary to communicate information about a neuron's activity. Some neurons communicate with simple resistive coupling, via channels that allow direction flow. However, for higher animals, most neurons in the central nervous system (CNS), communicate through chemical synapses: triggering the release of chemicals using the neural spike is called neurotransmitters

DOI: 10.4018/978-1-5225-5580-3.ch003

into the extracellular space. These neurotransmitters bind to ion channels in adjacent neurons, causing a brief ionic current to flow into the neuron. The resulting current flow in the recipient neuron will be depoloraizing, or hyperpolarizing and it depends on whether the neurotransmitter is excitatory or inhibitory, respectively.

It is very useful to have some knowledge of the way the biological nervous system is organized. Since the artificial neural network, draw much of their inspiration from the biological nervous system.

There is difference among the nervous system of creatures. Most living creatures, which have the ability to adapt to a changing environment, need a controlling unit, which is able to learn. Higher developed animals and humans use very complex networks of highly specialized neurons to perform this task.

In the living creatures, the brain is the control unit and it can be divided in different anatomic and functional sub-units. Each unit is responsible to do certain tasks like vision, hearing, motor and sensor control. The brain is connected by nerves to the sensors and actors in the rest of the body.

The brain consists of a very large number of neurons, about 10^{11} in average. These can be seen as the basic building bricks for the central nervous system (CNS). The neurons are interconnected at points called synapses. The massive number of highly interconnected simple units working in parallel, with an individual neuron receiving input from up to 10000 others causes the complexity of the brain (Bishop, C.M., 1995).

The neuron contains all structures of an animal cell. The complexity of the structure and of the processes in a simple cell is enormous. Even the cell body contains the organelles of the neuron and also the 'dendrites' originate there. These are thin and widely branching fibers, reaching out in different directions to make connections to a larger number of cells within the cluster. Input connections are made from the axons of other cells to the dendrites or directly to the body of the cell. These are known as axondentritic and axonsomatic synapses. There is only one axon per neuron. It is a single and long fiber, which transports the output signal of the cell as electrical impulses (action potential) along its length. The end of the axon may divide in many branches, which are then connected to other cells. The branches have the function to fan out the signal to many other inputs (Hassoun, M. H., 1995).

There are many different types of neuron cells found in the nervous system. The differences are due to their location and function. The neurons perform basically the following function: all the inputs to the cell, which may vary

by the strength of the connection or the frequency of the incoming signal, are summed up. The input sum is processed by a threshold function and produces an output signal. The processing time of about 1ms per cycle and transmission speed of the neurons of about 0.6 to 120 {ms} are comparingly slow to a modern computer.

The brain works in both a parallel and serial way. The parallel and serial nature of the brain is readily apparent from the physical anatomy of the nervous system. That there is serial and parallel processing involved can be easily seen from the time needed to perform tasks. For example a human can recognize the picture of another person in about 100 ms. Given the processing time of 1 ms for an individual neuron this implies that a certain number of neurons, but less than 100, are involved in serial; whereas the complexity of the task is evidence for a parallel processing, because a difficult recognition task cannot be performed by such a small number of neurons. This phenomenon is known as the 100-step-rule. Structurally the neuron can be divided in three

Figure 1. A biological neuron (Hassoun, M. H., 1995)

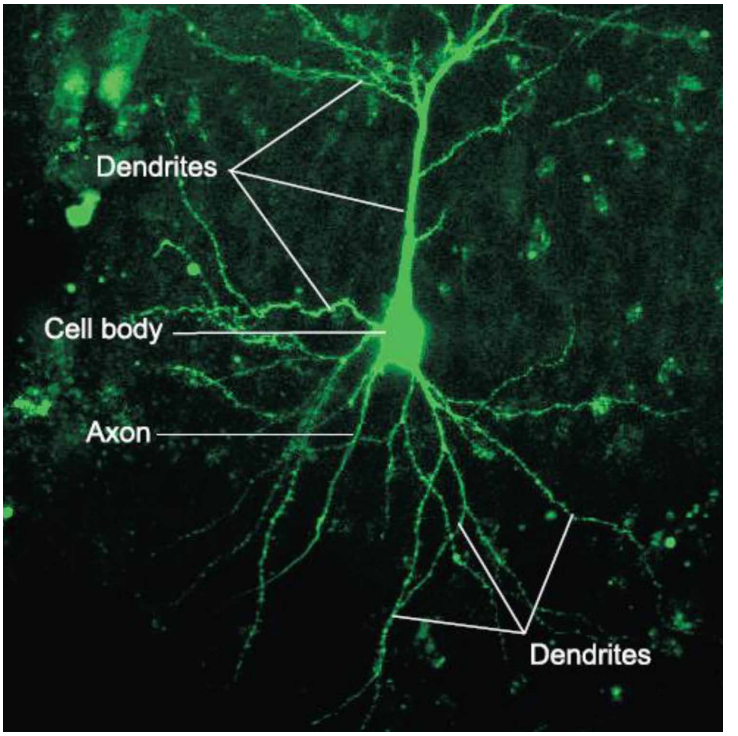
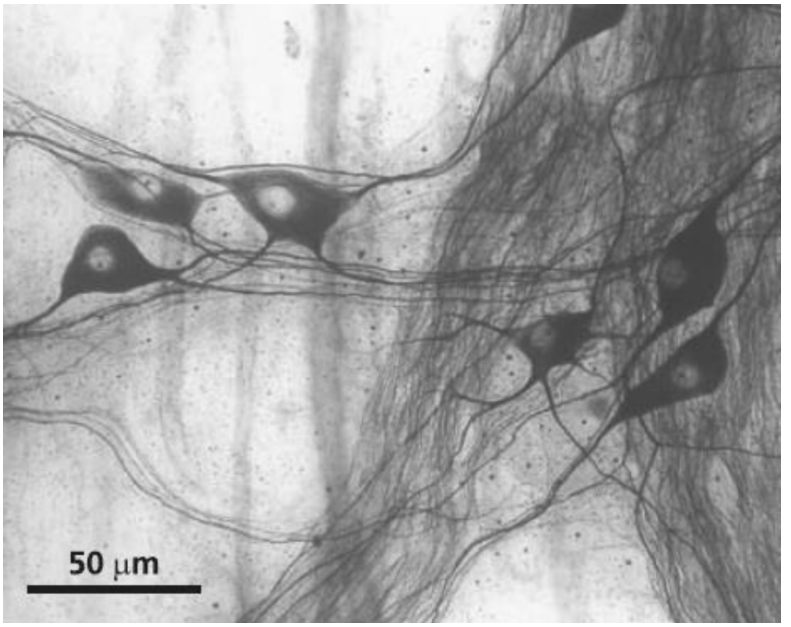


Figure 2. The main processes of the individual cells (Hassoun, M. H., 1995)



major parts: the cell body (soma), the dentrites, and the axon, see Figure 1 and Figure 2.

The cell body contains the organelles of the neuron and also the `dendrites' are originating there. These are thin and widely branching fibers, reaching out in different directions to make connections to a larger number of cells within the cluster. Input connections are made from the axons of other cells to the dendrites or directly to the body of the cell. These are known as axondentritic and axonsomatic synapses (Hassoun, M. H., 1995).

There is only one axon per neuron. It is a single and long fiber, which transports the output signal of the cell as electrical impulses (action potential) along its length. The end of the axon may divide in many branches, which are then connected to other cells. The branches have the function to fan out the signal to many other inputs.

There are many different types of neuron cells found in the nervous system. The differences are due to their location and function. The neurons perform basically the following function: all the inputs to the cell, which may vary by the strength of the connection or the frequency of the incoming signal, are summed up. The input sum is processed by a threshold function and produces an output signal. The processing time of about 1ms per cycle and

transmission speed of the neurons of about 0.6 to 120 {ms} are comparingly slow to a modern computer.

The neuron has four main regions to its structure. The cell body, or soma, has two offshoots from it, the dendrites, and the axon, which end in presynaptic terminals. The cell body is the heart of the cell, containing the nucleus and maintaining protein synthesis. A neuron may have many dendrites, which branch out in a treelike structure, and receive signals from other neurons. A neuron usually only has one axon which grows out from a part of the cell body called the axon hillock. The axon conducts electric signals generated at the axon hillock down its length. These electric signals are called action potentials. The other end of the axon may split into several branches, which end in a presynaptic terminal. Action potentials are the electric signals that neurons use to convey information to the brain. All these signals are identical.

Therefore, the brain determines what type of information is being received based on the path that the signal took. The brain analyzes the patterns of signals being sent and from that information it can interpret the type of information being received. Myelin is the fatty tissue that surrounds and insulates the axon. Often short axons do not need this insulation. There are uninsulated parts of the axon. These areas are called Nodes of Ranvier. At these nodes, the signal traveling down the axon is regenerated. This ensures that the signal traveling down the axon travels fast and remains constant (i.e. very short propagation delay and no weakening of the signal).

The synapse is the area of contact between two neurons. The neurons do not actually physically touch. They are separated by the synaptic cleft, and electric signals are sent through chemical 13 interaction. The neuron sending the signal is called the presynaptic cell and the neuron receiving the signal is called the postsynaptic cell. The signals are generated by the membrane potential, which is based on the differences in concentration of sodium and potassium ions inside and outside the cell membrane. Neurons can be classified by their number of processes (or appendages), or by their function. If they are classified by the number of processes, they fall into three categories. Unipolar neurons have a single process (dendrites and axon are located on the same stem), and are most common in invertebrates. In bipolar neurons, the dendrite and axon are the neuron's two separate processes. Bipolar neurons have a subclass called pseudo-bipolar neurons, which are used to send sensory information to the spinal cord. Finally, multipolar neurons are most common in mammals.

Examples of these neurons are spinal motor neurons, pyramidal cells and Purkinje cells (in the cerebellum). If classified by function, neurons again fall

into three separate categories. The first group is sensory, or afferent, neurons, which provide information for perception and motor coordination. The second group provides information (or instructions) to muscles and glands and is therefore called motor neurons. The last group, interneuronal, contains all other neurons and has two subclasses. One group called relay or projection interneurons have long axons and connect different parts of the brain. The other group called local interneurons are only used in local circuits.

Biological neural systems usually have a very high fault tolerance. Experiments with people with brain injuries have shown that damage of neurons up to a certain level does not necessarily influence the performance of the system, though tasks such as writing or speaking may have to be learned again. This can be regarded as re-training the network.

3.2 DEFINITION OF ARTIFICIAL NEURAL NETWORK

A generic artificial neural network can be defined as a computational system consisting of a set of highly interconnected processing elements, called neurons, which process information as a response to external stimuli. An artificial neuron is a simplistic representation that emulates the signal integration and threshold firing behavior of biological neurons by means of mathematical equations. Like their biological counterpart, artificial neurons are bound together by connections that determine the flow of information between peer neurons. Stimuli are transmitted from one processing element to another via synapses or interconnections, which can be excitatory or inhibitory. If the input to a neuron is excitatory, it is more likely that this neuron will transmit an excitatory signal to the other neurons connected to it. Whereas an inhibitory input will most likely be propagated as inhibitory.

Artificial neural networks are computational paradigms based on mathematical models that unlike traditional computing have a structure and operation that resembles that of the mammal brain. Artificial neural networks or neural networks for short, are also called connectionist systems, parallel distributed systems or adaptive systems, because they are composed by a series of interconnected processing elements that operate in parallel. Neural networks lack centralized control in the classical sense, since all the interconnected processing elements change or “adapt” simultaneously with the flow of information and adaptive rules. One of the original aims of

artificial neural networks (ANN) was to understand and shape the functional characteristics and computational properties of the brain when it performs cognitive processes such as sensorial perception, concept categorization, concept association and learning. However, today a great deal of effort is focused on the development of neural networks for applications such as pattern recognition and classification, data compression and optimization.

In the literature, a wide variety of definitions and explanations for the terms Artificial Neural Network and Neural Computing can be found. The following definitions are balanced towards computing but are nevertheless very comprehensive in my opinion and they offer a wide range of views of what an ANN is.

Complexification in artificial neural networks can prove to be as important, as it is in the development of natural neural systems. It is important in artificial development to unleash fitness potential otherwise left untouched and constrained by a fixed neural topology. Complexification in neural networks is a vital process in the development of the brain in any natural system. Complexification in human brains happens in several different ways, by growth, by pruning and by reorganization. The first form of complexification happens from before birth and goes on up to adulthood, as the brain is formed. During this period neurons and interconnections grow and hence complexifies the network. The second form of complexification happens through continuous pruning.

Connections between neurons have to be used for them not to fade away and eventually possibly disappear. This concept is called neural Darwinism, as it is similar to normal evolution; where the fittest, in this case connections, survive. The third form of complexification happens through reorganization. In some cases, for yet unknown reasons, connections detach themselves from neuron and reconnects to another. Mostly, reorganization in natural systems have a detrimental effect, but some might have unexpected positive effects (Hassoun, M. H., 1995).

The definition by Igor Aleksander is including a very wide range of methods and applications in the field of neural computing:

"Neural computing is the study of networks of adaptable nodes, which through a process of learning from task examples, store experimental knowledge and make it available for use." (Aleksander, I., and Morton H., 1995).

To simulate intelligent behavior the abilities of memorization and generalization are essential. These are basic properties of artificial neural

networks. The following definitions are according to the Collins English Dictionary:

- **To Memorize:** To commit to memory; learn so as to remember.
- **To Generalize:** To form general principles or conclusions from detailed facts, experience, etc.

Memorizing, given facts, is an obvious task in learning. This can be done by storing the input samples explicitly, or by identifying the concept behind the input data, and memorizing their general rules. The ability to identify the rules, to generalize, allows the system to make predictions on unknown data. Despite the strictly logical invalidity of this approach, the process of reasoning from specific samples to the general case can be observed in human learning.

The following description by Laurene Fausett of artificial neural networks includes only the connectionist research approach (Fausett, L., 1994).

"An artificial neural network is an information-processing system that has certain performance characteristics in common with biological neural networks. Artificial neural networks have been developed as generalizations of mathematical models of human cognition or neural biology, based on the assumption that:

- Information processing occurs at many simple elements called neurons.
- Signals are passed between neurons over connection links.
- Each connection link has an associated weight, which, in a typical neural net, multiplies the signal transmitted.
- Each neuron applies an activation function (usually nonlinear) to its net input (sum of weighted input signals) to determine its output signal."

Robbert L. Harvey focuses very much on the biological model. His definition excludes most parts of logical neural networks from the field of neural networks.

"An artificial neural network is a dynamical system with one-way interconnections. It carries out the processing by its response to inputs. The processing elements are nodes; the interconnections are directed links. Each processing element has a single output signal from which copies fan out."

The field of neural networks can be thought of as being related to artificial intelligence, machine learning, parallel processing, statistics, and other fields. The attraction of neural networks is that they are best suited to solving the

problems that are the most difficult to solve by traditional computational methods.

Partridge et al. (1996) listed several potentials of neural network over conventional computation and manual analysis:

- Implementation using data instead of possibly ill-defined rules.
- Noise and novel situations are handled automatically via data generalization.
- Predictability of future indicator values based on past data and trend recognition.
- Automated real-time analysis and diagnosis.
- Enables rapid identification and classification of input data.
- Eliminates error associated with human fatigue and habituation.

Generalization also removes the need to store a large number of input samples. Features common to the whole class need not to be repeated for each sample - instead the system needs only to remember which features are part of the sample. This can dramatically reduce the amount of memory needed, and produce a very efficient method of memorization.

Neural Network (NN) is a powerful AI technique that has the capability to learn a set of data and constructs weight matrixes to represent the learning patterns. NN is a network of many simple processors or units (Sarle, 1999). It simulates the function of human brain to perform tasks as human does. As an example, a study on approximation and classification in medicine with incremental neural network shows superior generalization performance compared with other classification models (Jankowski, 1999). NN has been employed in various medical applications such as coronary artery (Lippmann, 1995), Myocardial Infarction (Heden et al.,1996), cancer (Street et al., 1996; Karkanis et al., 1999), pneumonia (Caruana et al., 1996) and brain disorders (Pranckeviciene, 1999). In Karkanis et al (1999) NN was implemented as a hybrid with textual description method to detect abnormalities within the same images with high accuracy.

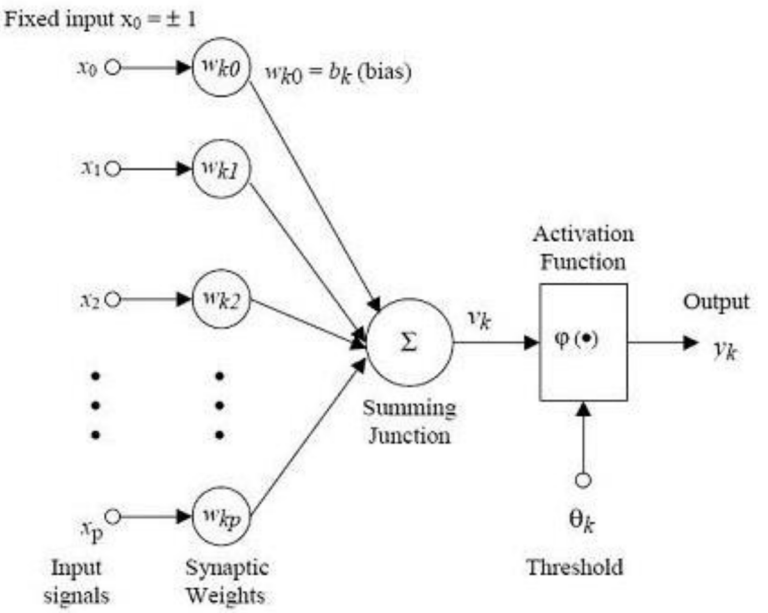
3.3 MATHEMATICAL MODEL OF NEURAL NETWORK

There are three basic components of importance for creating a functional model of the biological neuron. First, the synapses of the neuron are modeled as weights. The strength of the connection between an input and a neuron is

noted by the value of the weight. Negative weight values reflect inhibitory connections, while positive values designate excitatory connections. The next two components model the actual activity within the neuron cell. An adder sums up all the inputs modified by their respective weights. This activity is referred to as linear combination. Finally, an activation function controls the amplitude of the output of the neuron. An acceptable range of output is usually between 0 and 1, or -1 and 1 (Hassoun, M. H., 1995).

In the artificial neural network, each unit performs a relatively simple job: receive input from neighbors or external sources and use this to compute an output signal, which is propagated to other units. Apart from this processing, a second task is the adjustment of the weights. The system is inherently parallel in the sense that many units can carry out their computations at the same time. Within neural systems, it is useful to distinguish three types of units: input units, which receive data from outside the neural network, output units, which send data out of the neural network, and hidden units whose input and output signals remain within the neural network. During operation, units can be updated either synchronously or asynchronously. With synchronous updating, all units update their activation simultaneously; with asynchronous

Figure 3. The mathematical process of artificial neural network



updating, each unit has a (usually fixed) probability of updating its activation at a time t, and usually only one unit will be able to do this at a time. In some cases, the latter model has some advantages.

Mathematically, this process is described in Figure 3.

From this model, the interval activity of the neuron can be shown to be:

$$V_k = \sum_{j=1}^{p} w_{kj} xj \tag{1}$$

3.3.1 Activation Function

As mentioned previously, the activation function acts as a squashing function, such that the output of a neuron in a neural network is between certain values (usually 0 and 1, or -1 and 1). In general, there are three types of activation functions, denoted by Φ(.). First, there is the Threshold Function, which takes on a value of 0 if the summed input is less than a certain threshold value (v), and the value 1 if the summed input is greater than or equal to the threshold value.

The inputs received by a single processing element (depicted in Figure 3) can be represented as an input vector $A = (a_1, a_2, \ldots a_n)$, where a_i is the signal from the i^{th} input. A weight is associated with each connected pair of neurons. Hence weights connected to the j^{th} neuron can be represented as a weight vector of the form $W_j = (w_{1j}, w_{2j}, \ldots, w_{nj})$, where w_{ij} represents the weight associated to the connection between the processing element a_i, and the processing element a_j. A neuron contains a threshold value that regulates its action potential. While action potential of a neuron is determined by the weights associated with the neuron's inputs (Equation 1), a threshold v modulates the response of a neuron to a particular stimulus confining such response to a pre-defined range of values. Equation 2 defines the output y of a neuron as an activation function f of the weighted sum of n+1 inputs. These n+1 correspond to the n incoming signals. Equations (3), (4), and (5) are activation functions and are shown in Figures 4, 5, and 6, respectively.

$$y = f(v_k) \tag{2}$$

Figure 4. Step function

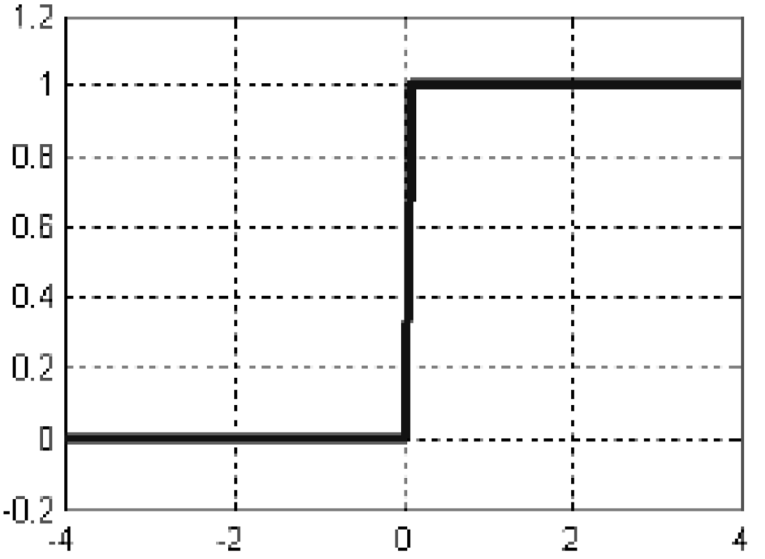


Figure 5. Sigmoid function

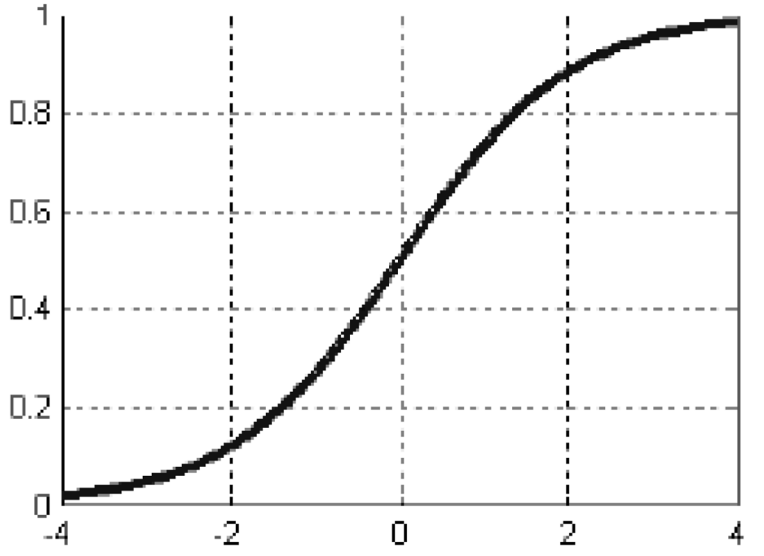
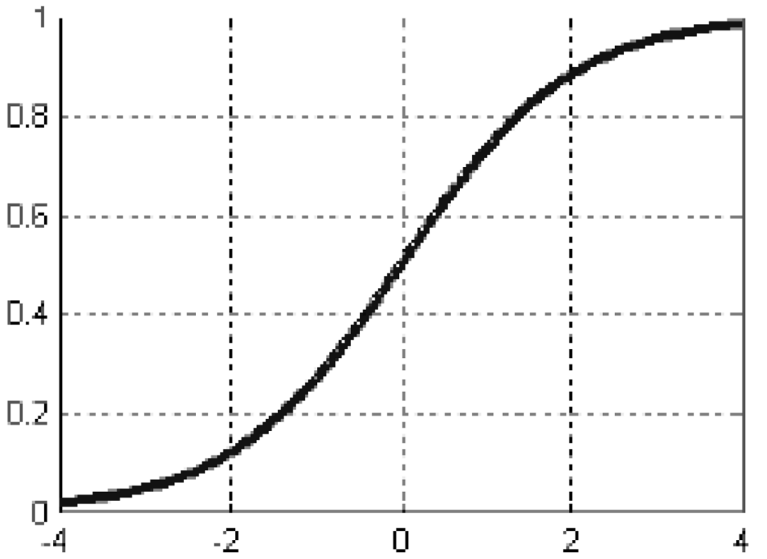


Figure 6. Hyperbolic tangent function



$$f(x) = \begin{cases} 1 \ if \ x > 0 \\ 0 \ if \ x \leq 0 \end{cases} \tag{3}$$

$$f(x) = \frac{1}{1 + e^{-x}} \tag{4}$$

$$f(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{3.5}$$

3.3.2 Training

Since the output(s) may not be what is expected, the weights may need to be altered. Some rule then needs to be used to determine how to alter the weights. There should also be a criterion to specify when the process of successive modification of weights ceases. This process of changing the weights, or rather, updating the weights, is called training. A network in which learning is employed is subjected to training. Training is an external process or regimen.

Learning is the desired process that takes place internal to the network. In the next chapter various types of neural network are explained (Hassoun, M. H., 1995). Following chapter explains about one of the optimization approaches that are used widely for solving different problems in the world and we apply them in combination with ANN to detect heart disorders.

3.4 SUMMARY

A generic artificial neural network can be defined as a computational system consisting of a set of highly interconnected processing elements, called neurons, which process information as a response to external stimuli. Artificial neural networks are computational paradigms based on mathematical models that unlike traditional computing have a structure and operation that resembles that of the mammal brain. In the artificial neural network, each unit performs a relatively simple job: receive input from neighbours or external sources and use this to compute an output signal, which is propagated to other units.

REFERENCES

Bello, R., & Verdegay, J. L. (2012). Rough sets in the soft computing environment. *Information Sciences*, *212*, 1–14. doi:10.1016/j.ins.2012.04.041

Chan, L.-W. (1999). Weighted least square ensemble networks. *Proceedings of the IJCNN*, *99*, 1393–1396.

Hansen, L. K., & Salamon, P. (1990). Neural network ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *12*(10), 993–1001. doi:10.1109/34.58871

Hassoun, M. H. (1995). *Fundamentals of artificial neural network*. Cambridge, MA: MIT.

Hornik, K. M., Stinchcombe, M., & White, H. (1989). Multilayer feedforward, networks are universal approximators. *Neural Networks*, *2*(5), 359–366. doi:10.1016/0893-6080(89)90020-8

Jorand, M. I., & Bishop, C. M. (1996). *Neural networks* (Technical Report No. A. I. Memo No. 1562). Artificial Intelligence Laboratory.

Krogh, A., & Vedelsby, J. (1995). Neural network ensembles, cross validation, and active learning. *Advances in Neural Information Processing Systems*, *7*, 231–238.

Krose, B., & Smagt, V. D. P. (1996). *An introduction to neural networks*. The University of Amsterdam.

Sarle, W. S. (1999). *Neural network FAQ, part 1 of 7: Introduction*. Retrieved from ftp://ftp.sas. com/pub/neurl/FAQ.html

Sharkey, S. W., Lesser, J. R., Zenovich, A. G., Maron, M. S., Lindberg, J., & Longe, T., F., & Maron, B. J. (2005). Acute and reversible cardiomyopathy provoked by stress in women from the United States. *Circulation, 111*, 472–479. doi:.CIR.0000153801.51470.EB10.1161/01

Sonoda, S., & Murata, N. (2013). *Nonparametric weight initialization of neural networks via integral representation*. Retrieved from http://arxiv.org/abs/1312.6461

Taniguchi, M., & Tresp, V. (1997). Averaging regularized estimator. *Neural Computation*. doi:10.1162/neco.1997.9.5.1163

Yao, X., & Liu, Y. (1997). A new evolutionary system for evolving artificial neural networks. *IEEE Transactions on Neural Networks*, *8*, 694–713. doi:10.1109/72.572107

Chapter 4
Swarm Optimization

ABSTRACT

In this chapter, one of the optimization algorithms based on swarm behaviour of agents in search space called swarm particle optimization (PSO) is introduced. Also, a description about how to use PSO for neural network training is provided.

4.1 SWARM INTELLIGENCE

Another part of the research of Huzly, H. (2006) is about Swarm Intelligence (SI). A part of her description in her research in provided here: "SI is the latest of an artificial intelligence technique based around the study of collective behaviour in decentralized and self-organized systems. The idea of SI came from systems found in nature, including ant colonies, bird flocking and animal herding that can be effectively applied to computationally intelligent system. SI systems are typically made up from a population of agents interacting locally with one another and with their environment and local interactions between such nodes often lead to the emergence of a global behaviour. There are two major techniques in SI, which are the Ant Colony Optimization (ACO) and Particle Swarm Optimization (PSO). The ACO algorithm is a probabilistic technique for solving computational problems to finding good paths through graphs. They are inspired by the behaviour of ants in finding paths from the colony to food. While PSO (which is the focus of this project) is a technique where all the particles (or solutions) move to get better results. PSO is a new branch of the soft computing paradigms called evolutionary algorithms (EA)

DOI: 10.4018/978-1-5225-5580-3.ch004

(Ismail, W., and shamsuddin, S. M., 2008). EA includes genetic algorithms (GA), evolutionary programming (EP), evolutionary strategies (ES) and genetic programming (GP). Before PSO, the most popular technique in evolutionary computing is Genetic Algorithm (GA). GA is widely used to determine BP learning parameters and weight optimization to make the convergence rate faster and avoid from being trapped in the local minima."

4.2 PARTICLE SWARM OPTIMIZATION

The original PSO algorithm is discovered through simplified social model simulation (Shi, 2004). PSO is a simple concept adapted from nature decentralized and self-organized systems such as choreography of a bird flock and fishing schooling. PSO is a population-based algorithm in which individual particles work together to solve a given problem. In PSO, physical position is not an important factor. The Population (or swarm) and the member called particle is initialized by assigning random positions and velocities and potential solutions are then flown through the hyperspace. The particles learn over time in response to their own experience and the experience of the other particles in their group (Ferguson, 2004). As mentioned before, PSO was introduced by Kennedy and Eberhart.

However in 1995, nowadays this concept has been explored by many other researchers around the globe and has been applied in many applications. Below are some application examples using PSO for optimization:

1. Application of Particle Swarm Optimization to design the electromagnetic absorbers by Suomin Cui* and Daniel S. Weile. (2005) Dept. of Electrical & Computer Engineering, University of Delaware. The synchronous PSO was applied to optimize multilayer coatings and polygonal absorbers for wide band frequency and/or wide incident range.
2. Human Tremor Analysis Using Particle Swarm Optimization by Russell C., Eberhart, and Xiaohui Hu (1999) where they present methods for the analysis of human tremor using particle swarm optimization. Two forms of human tremor are addressed which are essential tremor and Parkinson's disease.
3. Particle Swarm Optimization methods for pattern recognition and image processing by Mahamed G. H. Omran (2004) where PSO has been used to classify objects into different categories.

According to Eberhart *et al.* (2001), each particle keeps track of its best fitness position in hyperspace that has achieved so far. This value is called personal best or pbest. The overall best value obtained by so far by any particle in the population is called global best or gbest. During each epoch (or iteration) every particle is accelerated towards its own personal best as well as in the direction of the global best position. This is achieved by calculating a new velocity term for each particle based on the distance from its personal best, as well as its distance from the global best position. These two components ('personal' and 'global' velocities) are then randomly weighted to produce the new velocity value for this particle, which will in turn affect the next position of the particle during the next epoch. (Van den Bergh *et al*., 2000). Figure 1 shows the basic PSO procedure.

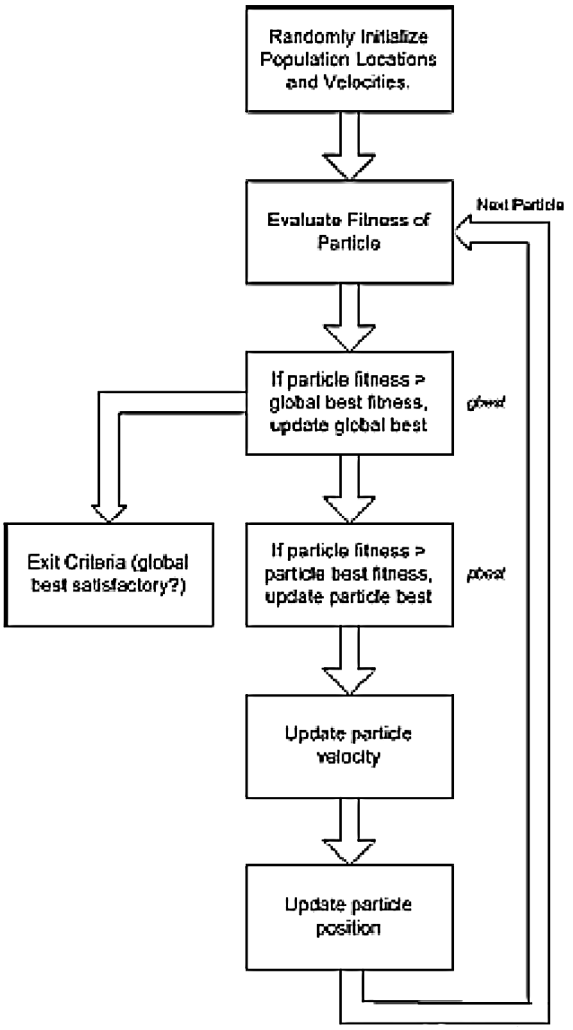
PSO is an optimization algorithm that using only primitive mathematic calculation. The advantage of the PSO over many of the other optimization algorithm is its relative simplicity (Van den Bergh *et al.,* 2000). According to Jones (2005), there are only two equations in PSO, the movement equation (Equation 1) and velocity update equation (2). The movement equation provides for the actual movement of the particles using their specific vector velocity while the velocity updates equation provides for velocity vector adjustment given the two competing forces (gbest and pbest).

$$x_n = x_n + v_n * \Delta t \tag{1}$$

$$v_n = v_n + c1 * rand() * (g_{best,n} - x_n) + c2 * rand() * (p_{best,n} - x_n) \tag{2}$$

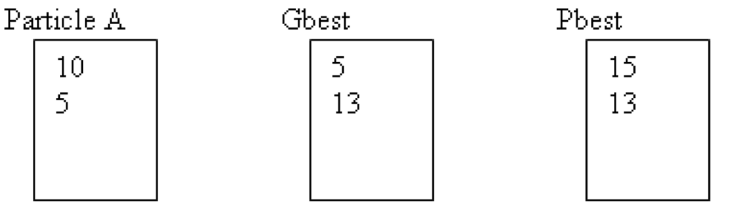
Equation 1 is performed for each element of the position (x) and velocity (v) vector. The Δt parameter (which is usually set to 1.0) defines the discrete time interval over which the particle will move. The result is a new position for the particle. In equation 2, it subtracts the dimensional element from the dimension from the best vector and then multiply this by a random number (between 0.0 and 1.0) and an acceleration constant (c_1 and c_2). The sum of these products is then added to the velocity for the given dimension of the vector. This process is performed for each element of the velocity vector. The random numbers provide an amount of randomness in the path to help the particle move throughout the solution space. The c_1 and c_2 acceleration constant provide some control to the equation to define which should be

Figure 1. Basic PSO procedure



given more emphasis on the path (global or personal best). The example below demonstrates how Particle A moves to the solution (gbest) in 2D space (which mean particle with two values). In this example, $c_1 = 1.0$, $c_2 = 0.5$ and $\Delta t = 1.0$. Because of the value of C1 is higher than C2, Particle A will give more emphasis to the global solution. Current particles position as shown in Figure 2.

Figure 2. Particles position



Assume that Particle A velocities values that calculated in previous iteration is Pv = (0, 1). First, the velocity vector must be updated for the current iteration using Equation 2.

Particle A first position (value = 10)

$$v_n = v_n + c1 * rand() * (g_{best,n} - x_n) + c2 * rand() * (p_{best,n} - x_n)$$

v_x=0+1.0*0.35*(5-10)+0.5*0.1*(15-10)

v_x=0+0.35*(-5)+0.05*(5)

v_x=0+(-1.75)+0.25

v_x=-1.5

Particle A second position (value = 5)

$$v_n = v_n + c1 * rand() * (g_{best,n} - x_n) + c2 * rand() * (p_{best,n} - x_n)$$

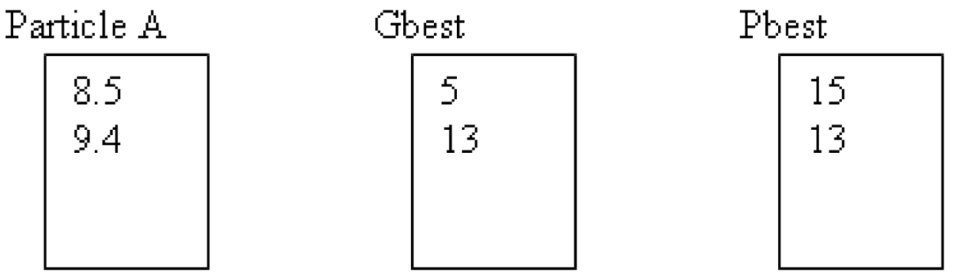
v_y=1+1.0*0.2*(13-5)+0.5*0.45*(13-5)

v_y=1+0.2*+0.225*(8)

v_y=1+1.6+1.8

v_y=4.4

Figure 3. New particles position



Now, current velocity value is Pv = (-1.5, 4.4). Than apply this new velocities to the particle positions using equation 2j.

$Position_x = position_x + v_x * \Delta t$

$Position_x = 10 + (-1.5) * 1.0$

$Position_x = 8.5$

$Position_y = position_y + v_y * \Delta t$

$Position_y = 5 + (4.4) * 1.0$

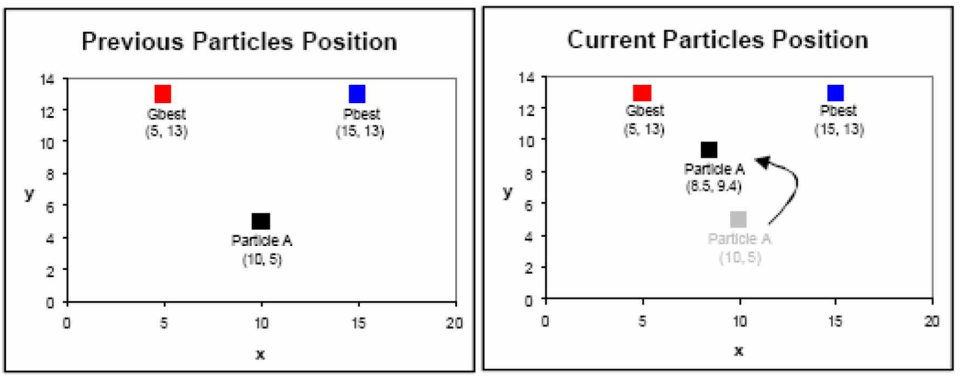
$Position_y = 9.4$

Based on the above calculations, the new position for Particle A as shown in Figure 3.

Figure 4 shows how Particle A moves in 2D space.

For a neural network implementation based on Al-kazemi *et al.* (2002), the position of each particle in swarm represents a set of weight for the current epoch or iteration. The dimensionality of each particle is the number of weights associated with the network. The particle moves within the weight space attempting to minimize learning error (or Mean Squared Error-MSE or Sum of Squared Error-SSE). Changing the position mean updating the weight of the network in order to reduce the error of the current epoch. In each epoch, all the particles update their position by calculating the new velocity, which they use to move to the new position. The new position is a set of new weights used to obtain the new error. For PSO, the new weights are adapted

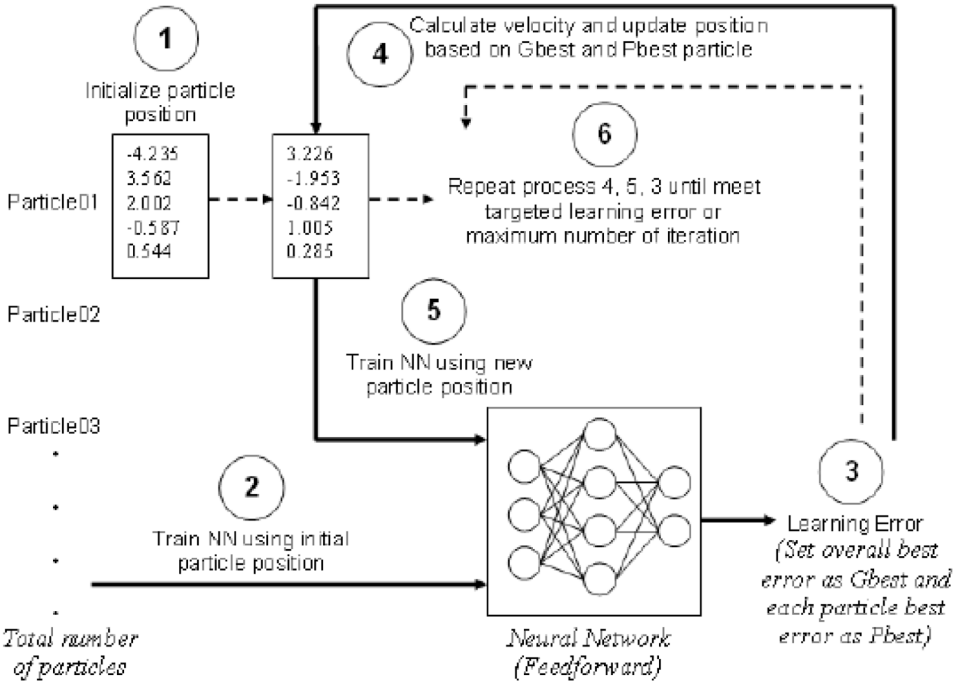
Figure 4. Particle A movement



even though no improvement is observed. This process is repeated for all the particles. The particle with the lowest error is considered as the global best particle so far. The training process continues until satisfactory error is achieved by the best particle or computational limits are exceeded. When the training ends, the weights are used to calculate the classification error for the training patterns. The same set of weights is used then to test the network using the test patterns. There is no backpropagation concept in PSONN where the feedforward NN produced the learning error (particle fitness) based on set of weight and bias (PSO positions). The pbest value (each particle's lowest learning error so far) and gbest value (lowest learning error found in entire learning process so far) are applied to the velocity update equation (2) to produce a value for positions adjustment to the best solution or targeted learning error. The new sets of positions (NN weight and bias) are produced by adding the calculated velocity value (equation 2) to the current position value using movement equation (1). Than that new set of positions are uses for producing new learning error (particle fitness) in feedforward NN. This process is repeated until the stop conditions are met (Minimum learning error or maximum number of iteration). In this study, the classification output has been written to the text file based on gbest position value. Compared to BP, learning error is calculated in feedforward NN starting from input nodes to hidden nodes and output nodes, then NN make a backward pass from output nodes to hidden nodes and input nodes to produce new set of weights. The summary on PSONN learning process is shown in Figure 5.

PSONN program in this study has been developed based on PSO program for Sombrero function optimization. The particle position represents two-dimensional (2D) vector of X and Y values in Sombrero function. The

Figure 5. PSONN learning process



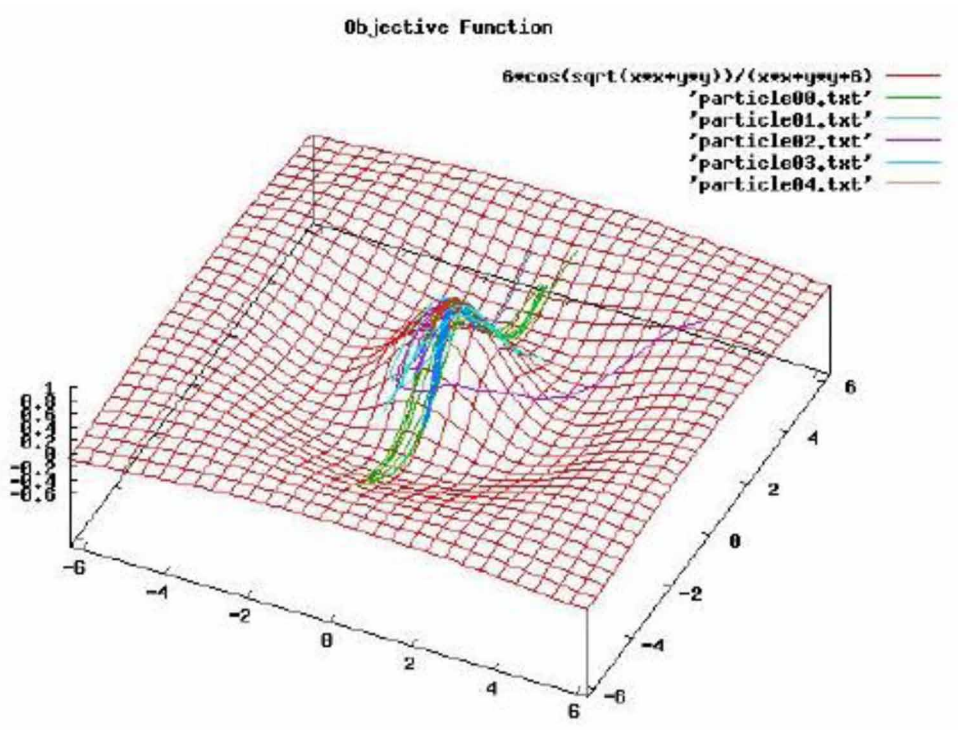
objective is to reach the value of 1 based on value of X and Y in Sombrero equation as shown in 2l and the goal for the PSO is to maximize the function.

$$z = 6 * \cos\left(\frac{\sqrt{x * x + y * y}}{x * x + y * y + 6}\right) \tag{3}$$

Figure 6 shows the how 5 particles that have been used to solve the Sombrero function move or fly to the solution in 2D problem. The fitness of the particle on the Sombrero is represented by Z-axis.

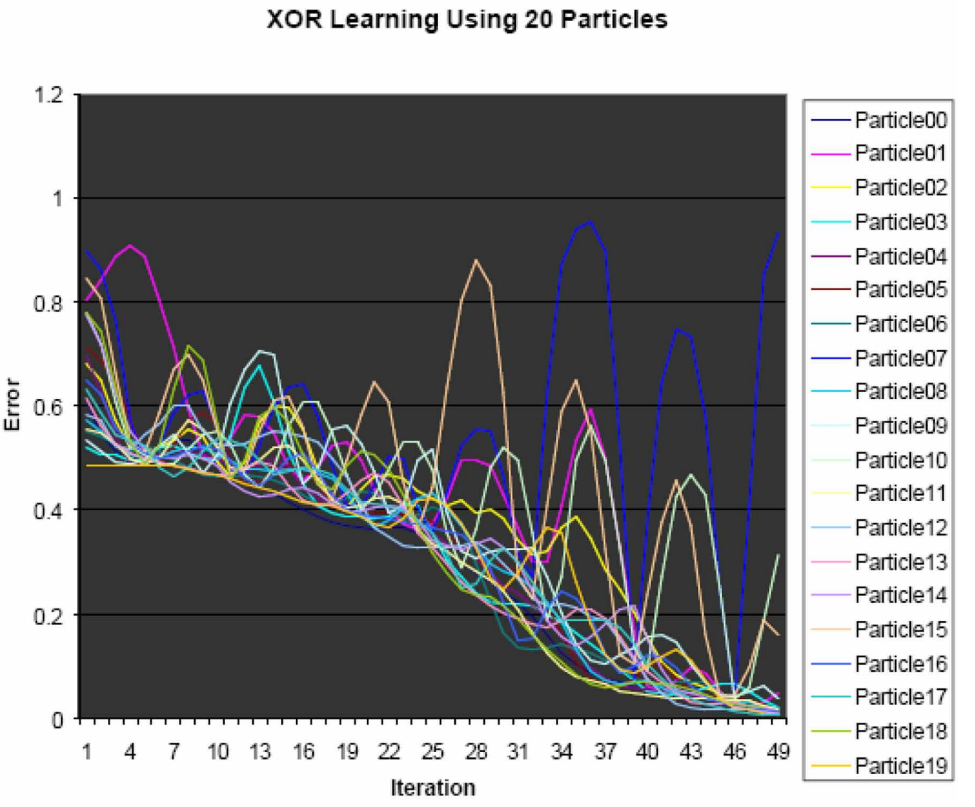
In this chapter, PSO is applied to neural network. To explain how PSONN works, Figure 7 shows 20 particles have been used to solve the XOR problems. The optimization objective is to get the Mean Squared Error (MSE) or learning error less than 0.005. The figure shows that at the beginning, all particles with random weight (position) have different MSE. During the learning, all particles seem to fly or move together to get the lowest MSE based on the best MSE found by one particle in every iteration. In each particle, the

Figure 6. Particle movement in Sombrero function optimization



best MSE found by every particle is called pbest fitness while the best MSE found from all particles is called gbest fitness and gbest is a solution for the optimization problem. Even though there are several particles that flown or move far away from the group of particles, but it managed to follow back the group of particles and try to get the best MSE. This is similar to the bird flocking concept where if there any bird that misses the group, it will try to fly back to their group and fly together to the targeted destination. This is why the particle also known as solution because of each particle represents the solution for the problem that needs to be optimized. In this example, Particle17 rich the targeted MSE first with MSE = 0.00472826 in 11 seconds at 49 iterations.

Figure 7. Particle movement in XOR problems



4.3 SUMMARY

There are optimization algorithms for increasing the performance and speed of training. One algorithm that usually is used for optimization is Swarm Intelligence (SI). In this chapter, PSO is applied to neural network. Also, it is conducted to prove the effectiveness PSO-based neural network and compared to GA-based neural network based on several universal data for classification problem.

REFERENCES

Bergh, V. D. F. (1999). *Particle swarm weight initialization in multi-layer perceptron artificial neural networks*. Paper accepted for ICAI, Durban, South Africa.

Bergh, V. D. F., & Engelbrecht, A. P. (2000). Cooperative learning in neural networks using particle swarm optimizers. *South African Computer Journal*, *26*, 84–90.

Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford, UK: Oxford University Press.

Bonabeau, D., & Theraulaz, G. (1999). *Swarm intelligence: From natural to artificial system*. New York: Oxford University Press.

Chun-an, L. (2008). New dynamic constrained optimization PSO algorithm. In *Proc. Fourth International Conference on Natural Computation* (pp. 650-653). Academic Press.

Devi, S., Jadhav, D. G., & Pattnaik, S. S. (2011). *PSO based memetic algorithm for unimodal and multimodal function optimization. In Swarm, Evolutionary, and Memetic Computing, LNCS, 7076* (pp. 127–134). Berlin: Springer.

Dorigo, M., Maniezzo, V., & Colorni, A. (1996). The ant system: Optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man, and Cybernetics*, *26*(1), 29–41. doi:10.1109/3477.484436 PMID:18263004

Eberhart, R., & Shi, Y. (2001). Particle swarm optimization: Developments, application and resources. In *Proceedings of the 2001 Congress on Evolutionary Computation* (pp. 81-86). IEEE. doi:10.1109/CEC.2001.934374

Farmer, J. D., Packard, N. H., & Perelson, A. S. (1986). The immune system, adaptation and machine learning. *Physica D. Nonlinear Phenomena*, *22*(1-3), 187–204. doi:10.1016/0167-2789(86)90240-X

Ferguson, D. (2004). *Particle swarm*. University of Victoria.

Flake, G. (1999). *The computational beauty of nature*. Cambridge, MA: MIT Press.

Fnaiech, F., Abid, S., Ellala, N., & Cheriet, M. (2002). A comparative study of fast neural network learning algorithms. In *Proceedings of IEEE International Conference on Systems, Man and Cybernetics (IEEE SMC)* (pp. 24-29). IEEE. doi:10.1109/ICSMC.2002.1173379

Formato, R. A. (2008). Central force optimization: A new nature inspired computational framework for multidimensional search and optimization. *Studies in Computational Intelligence*, *129*, 221–238. doi:10.1007/978-3-540-78987-1_21

Cui, S., & Weile, D. S. (2005, November). Application of a parallel particle swarm optimization scheme to the design of electromagnetic absorbers. *IEEE Transactions on Antennas and Propagation*, *53*(11), 3616–3624. doi:10.1109/TAP.2005.858866

Eberhart, R. C., & Hu, X. (1999). *Human Tremor Analysis Using Particle Swarm Optimization*. Retrieved from www.swarmintelligence.org/papers/CEC1999Human.pdf

Omran, M. G. H. (2004). *Particle Swarm Optimization Methods for Pattern Recognition and Image Processing*. Retrieved from https://pdfs.semanticscholar.org/f55a/20f7f7a6ba00ecbd958be6cfb2c60324fe06.pdf

Chapter 5
Feature Extraction

ABSTRACT

This chapter is an illustration of feature extraction for working with large datasets. The basic definition of feature extraction, selection of effective features, and the existing problems and solutions are provided. How feature extraction maps the high dimensional space to smaller space is explained.

5.1 INTRODUCTION

Feature extraction is an attribute reduction process. Unlike feature selection, which ranks the existing attributes according to their predictive significance, feature extraction actually transforms the attributes. The transformed attributes, or features, are linear combinations of the original attributes.

Digital libraries are places for handling a vast amount of data and information. This provides the ability to access and interact to a lot of documents in the form of electronic version. Schatz (1997) has a definition for digital libraries: "A digital library enables users to interact effectively with information distributed across a network".

A digital library works as a network information system to support some tasks such as search and display of an item in a database. Once comfortable with the new tools, they demand new materials to be available in digital libraries. Any task in the digital library requires the digital representation of the data. Extending a digital library is an easy task, because of faster and cheaper process than the physical library. Obviously, this increase in the amount of information has a strong impact on the supporting software.

DOI: 10.4018/978-1-5225-5580-3.ch005

Three are multimedia content in the electronic version of text data, which is an issue. To obtain the image in the document, one can provide the query and ask for all the electronic documents that contain similar pictures. Digital images and any other multimedia data for that matter are *complex data*. Computers have the ability to represent and manipulate the digital data. However, decoding the contents is an issue for researching (Manolescu, 2000).

5.1.1 Problem

The question is that how a software can handle the requirements related to the applications to deal with large amounts of information, similarity searching and complex data.

5.1.2 Forces

- Difficulties in information retrieval systems with large amounts of data.
- Difficulties in similarity searching.
- Difficulties in multimedia databases containing digital representations of acoustic and visual data.
- Difficulties in information retrieval systems for fast responding time.

5.1.3 Mapping the Data

The solution to overcome the difficulties is to work with a simpler presentation of the data. The representation contains the unique form for each data. There are functions for these computations. It means that the function maps the data from the problem space into a feature space. For this reason, it is also called a "feature extraction function". Feature extraction is mapping the data from a larger problem space into a smaller feature space. Therefore, feature extraction enables scalable solutions for problems that deal with large amounts of information (Manolescu, 2000).

Feature extraction has some limitation for similarity searching problem. Any data items are similar as long as their representation in the feature space are similar. But this is not necessarily how humans perceive things. Two reasons for this are the following:

1. The problem space may be ambiguous. Text is a notorious example. Humans handle this problem by using the surrounding text to establish a context.
2. The feature extraction function is non-injective. Distinct points from problem space can map into the same point in feature space. More on this issue later.

Using the mapping technique a complex domain is transformed into a simpler one. An example is the operational method for the solution of differential equations (Bronshtein & Semendyayev, 1997). The method consists in mapping the data from a differential equation, by means of an integral transformation, to a transformed equation. The transformed equation is easier to solve than its differential counterpart. The Laplace transform and the z transform are two possible way for transforming.

Computers have the ability to manipulate the complex information mapping from the problem space to feature space by provided softwares. Digital images are one example. Current image databases employ this pattern to obtain simplified representations for images. Unlike the typical domain mappings from mathematics, these simplified representations lose information. They consider only the *most significant* image features, e.g., the low-frequency coefficients of the Fourier transform. Common features for images are colour histograms, textures, shapes or a combination of these. This way, feature extraction enables software systems to process different types of complex information without "understanding" the contents (Manolescu, 2000).

When mapping from a large problem space, feature extraction considers only a few "significant" features in the feature space, discarding the rest. This truncation yields a non-injective mapping. For example, two documents can map into the same point in keyword space. However, this does not mean they are identical.

Since the function is not injective, there is no inverse mapping. Several points in problem space can map into a single point in the feature space. This property affects all applications that employ this pattern to provide answers to queries. The solution is to add a post-processing step that filters out the "false alarms." Since the typical number of false alarms is small, the post-processing step usually performs a sequential search to eliminate them. Besides post-processing, this pattern requires two additional stages:

1. Feature extraction works with the features of the working set of items (document, images, etc.). Whenever a new item is added to the data store, the system computes its features (i.e., coordinates in the feature space). Therefore, each *insertion* needs this extra step.
2. Another operation that changes is *query processing*. The fundamental idea of feature extraction is to perform all computations in a smaller, simpler space. Processing then takes place in this space. Consequently, answering a query requires computing its representation in the feature space as well (Manolescu, 2000).

Information retrieval (IR) is one of the domains that employs feature extraction extensively. IR has expanded into fields such as office automation, genome databases, fingerprint identification, medical image management, data mining and multimedia. In many of these applications the objective is to minimize response times for different sorts of queries. Performance depends on how fast the system performs searches in the multidimensional feature space. Therefore, the choice of a spatial access method (SAM) is critical. However, this may be challenging. Good unidimensional indexing methods scale exponentially for high dimensionalities, eventually reducing to sequential scanning (Agrawal et al., 1993). Therefore, they apply only when a small number of dimensions is sufficient to differentiate between data items. The next section ("Design decisions") provides a few alternatives that work well for a larger number of dimensions.

5.1.4 Design Decisions

Applying the feature extraction pattern involves three design decisions. This section covers these decisions and discusses some potential choices.

The creative part of this pattern is obtaining a suitable feature extraction function. This is also the hard part of the pattern. One of the important requirements for the domains that employ feature extraction is correctness. A query should return all the qualifying information, without any "misses." The "false alarms" due to the non-injective mapping are less of a problem; post-processing removes them. However, a formal proof is required to demonstrate correctness. Alternatively, a domain-specific algorithm may automatically construct a correct feature-extraction function for a given problem. For example, (Faloutsos & Lin, 1995) describes such an algorithm for indexing, data-mining and visualization of traditional and multimedia

datasets. Obviously, the feature extraction function is domain- and problem-dependent.

After finding a feature extraction function, we must decide which features to consider further. As explained before, not all features are used. For example, systems that use DFT keep only a few low frequency coefficients. This "lossy" part of the pattern ensures that feature space is smaller than problem space. Deciding on the number of features involves a trade-off between accuracy and speed. At one extreme, the system is "lossless" and keeps all features. This ensures no false alarms. However, searching a large (feature) space is what the pattern is trying to avoid. At the other extreme, only one feature is used. In this case, the degenerate search in the feature space is fast—it simply returns everything. Post-processing takes a long time though, since it filters all data items. Therefore, the number of features determines the balance between the searching time in the feature space and the post-processing time.

The third part of this pattern is choosing a suitable spatial access method. The choice depends on the number of features—dimensions of the feature space. Many methods are available for indexing low dimensionality domains—for example, hash tables or B-tree variants. However, as the number of dimensions grows, they degenerate into sequential scanning. R-tree variants (e.g., R*-trees (Beckmann et al, 1990), and Strees (David, 1996) offer good performance for a larger number of dimensions. To summarize, the feature extraction pattern involves three design decisions:

- Determine the feature extraction function. This is the most challenging part of the pattern.
- Decide what features to consider. This decision determines the balance between the search time and the post-processing time.
- Choose a spatial access method. This determines how fast the system can search the feature space.

5.1.5 Consequences

Feature extraction provides solutions for important information retrieval problems. First, it enables scalable solutions for systems that deal with large amounts of information. Second, it provides a natural and low overhead solution for similarity search. And finally, it enables software to process different types of complex information without "understanding" its contents (Manolescu, 2000).

The feature extraction pattern has the following benefits (✓) and liabilities (✗):

✓ It can manage large amounts of data. Compared to sequential scanning, applications using this pattern obtain an increasingly better performance as the volume of data increases (Agrawal R., 1993).
✓ Similarity searching corresponds to vector operations in feature space. These have low computational overhead and rank the results.
✓ Software can manipulate complex information without having to decode its semantics. This is key for implementing multimedia databases.
✓ Users can easily refine queries. Once results are available, they mark only those that are relevant. The system adjusts the original query and performs a new search. If the user's feedback is consistent, such queries converge in a few iterations. This mode of operation is also known as "relevance feedback" (Salton & Buckley, 1998). In the feature space, relevance feedback consists of adding the selected vectors to the query vector.
✗ It is hard to determine feature extraction functions. This is often the subject of doctoral dissertations or even careers.
✗ Efficient search in the feature space requires spatial access methods. Not all good indexing methods scale well with the number of dimensions. Obtaining an efficient and scalable multidimensional index structure is difficult.
✗ Inserting new items and answering queries require additional processing. The architect has to determine the right balance between the number of features and the post-processing time.
✗ The features require an additional data store. Systems that use feature extraction use two data stores.

The "problem space" data store holds the domain-specific entities, e.g., documents, images, etc. Likewise, the "feature space" data store holds the corresponding features.

5.2 FEATURE SELECTION

We are decomposing the problem of feature extraction in two steps: feature construction, briefly reviewed in the previous section, and feature selection,

to which we are now directing our attention. Although feature selection is primarily performed to select relevant and informative features, it can have other motivations, including:

1. General data reduction, to limit storage requirements and increase algorithm speed;
2. Feature set reduction, to save resources in the next round of data collection or during utilization;
3. Performance improvement, to gain in predictive accuracy;
4. Data understanding, to gain knowledge about the process that generated the data or simply visualize the data.

5.3 SUMMARY

Feature extraction is an attribute reduction process. Unlike feature selection, which ranks the existing attributes according to their predictive significance, feature extraction actually transforms the attributes. Typically, feature extraction maps from a *larger* problem space into a *smaller* feature space. Data are represented by a fixed number of features, which can be binary, categorical or continuous. Feature is synonymous of input variable or attribute. Finding a good data representation is very domain specific and related to available measurements.

REFERENCES

Al-kazemi, B., & Mohan, C. K. (2002). Training feedforward neural network using multi-phase particle swarm optimization. *Proc 9th Intl Conf Neural Infor Process*, (2), 2615-2619. 10.1109/ICONIP.2002.1201969

Bergh, F. V. D. (2001). *An analysis of particle swarm optimizers* (Thesis). University of Pretoria, Pretoria, South Africa.

Bishop, C. M. (2007). *Pattern recognition and machine Learning*. Information Science and Statistics.

Blondin, J. (2009). *Particle Swarm Optimization: A Tutorial*. Amsterdam University.

Castro, B., Kogan, D., & Geva, A. B. (2000). ECG feature extraction using optimal mother wavelet. *The 21st IEEE Convention of the Electrical and Electronic Engineers*, 346-350. 10.1109/EEEI.2000.924422

Ebrahim, Z. A., Khazaee, A., & Ranaee, V. (2010). Classification of the electrocardiogram signals using supervised classifiers and efficient features. *Comp Meth Prog Biomed*, (99), 179-194.

Fredric, M. H., & Soowhan, H. (1996). Classification of cardial arrhythmias. *IEEE Transactions on Biomedical Engineering*, *43*(4), 425–429. doi:10.1109/10.486263 PMID:8626192

Gacek, A., & Pedrycz, W. (2012). *ECG Signal Processing, Classification and Interpretation*. New York: Springer. doi:10.1007/978-0-85729-868-3

Karakülah, G., Dicle, O., & Koşaner, O. (2014). Computer based extraction of phenoptypic features of human congenital anomalies from the digital literature with natural language processing techniques. *Studies in Health Technology and Informatics*, *205*, 570–574. PMID:25160250

Manolescu, D. A. (2000). *Feature Extraction—A Pattern for Information Retrieval*. Retrieved from https://micro-workflow.com/PDF/plop98.pdf

Chapter 6
Literature Review

ABSTRACT

This chapter introduces different resources about noise in heart signals. It also provides a short explanation about artificial neural network (ANN), particle swarm optimization (PSO), and presents some of the previous studies related to heart signal noise removal, intelligent methods for detection of disorders, and feature extraction.

6.1 INTRODUCTION

In this chapter, various types of noise that interfere with the ECG signal are discussed in Section 6.1. In Sections 6.2 and 6.3, the basic fundamentals of artificial neural networks (ANN) and swarm intelligence (SI) are explored. In section 6.4, a literature review on previous related studies is provided. Finally the last section summarises the chapter.

6.2 NOISE IN ECG SIGNAL

There are various sources of noise including imperfect contact of electrodes to the body, machine malfunction, electrical noise from elsewhere in the body, respiration and muscle contractions, which corrupts the ECG signal (Poungponsri and Yu, 2009). The produced noise consists of low-frequency components and high-frequency components that cause baseline wander

DOI: 10.4018/978-1-5225-5580-3.ch006

and powerline interference, respectively (Jesmin et al., 2011). Below is a description for different types of noise in ECG signals.

6.2.1 Powerline Interference

Powerline interference is the 50 or 60 Hz pickup that is caused by improper grounding (Behbahani, 2007). This is due to an impulse or spike at 60 Hz/50 Hz harmonics which can be removed using a 60 Hz notch filter. Figure 1 shows the powerline interference that has corrupted the ECG.

6.2.2 Electrode Contact Noise

Another type of noise that is caused by poor contact between the electrodes and skin is transient interference that effectively reduces the accuracy of the measurement device. Figure 2 shows the false beat in the signal, which is because of the imperfect electrode contact. The contact may be loose

Figure 1. 60 Hz powerline interference

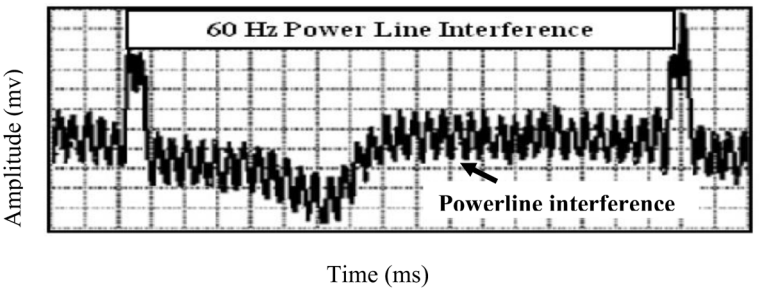
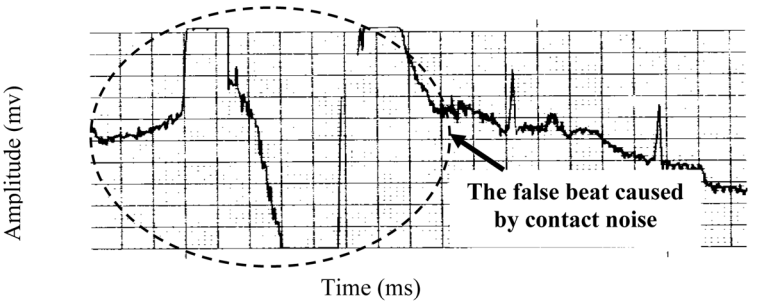


Figure 2. Electrode contact noise



permanently, or intermittently. This loose contact may be as the result of the movement and vibration of the electrodes on the skin (Pomi and Olivera, 2006) and can result in large artefacts.

6.2.3 Motion Artefact

Motion artefacts are transient baseline change due to the electrode skin impedance with electrode motion. This motion generates the larger amplitude in the ECG signal. The motion artefact sometimes has the peak amplitude of 500 percent of the ECG. Its duration is about 100 – 500 ms (Pomi and Olivera, 2006). This noise can be removed using an adaptive filter. Figure 3 shows the motion artefact that cause similar amplitude for P, QRS and T in waveform.

Skin stretching causes changing in the amplitude of P, QRS, and T wave, which is because of some mechanical disturbances of the distribution of electrode-electrolyte interface. It causes higher amplitude for P, QRS, and T wave than the normal.

6.2.4 Baseline Drift

Respiration and movements of body cause baseline drift in the ECG signals that can vary the ECG amplitude about 15 percent. However, inaccurate baseline drift attenuation can cause clinical information distortion for instance ST segment distortion, because of the ECG signal overlaps in the spectrum of low frequency components and baseline drift (Raj and Venkateswarlu, 2011). Figure 4 shows the baseline drift on ECG signal.

Figure 3. Motion artefact

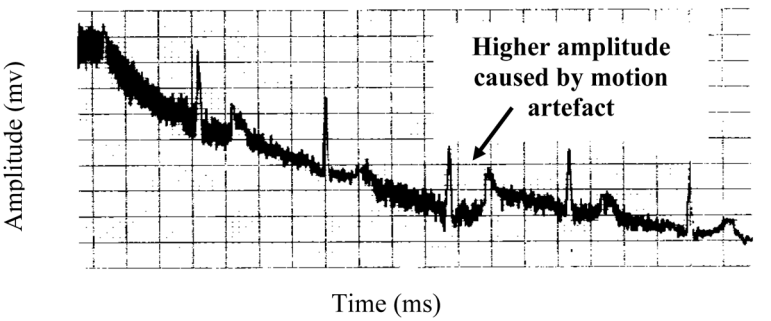
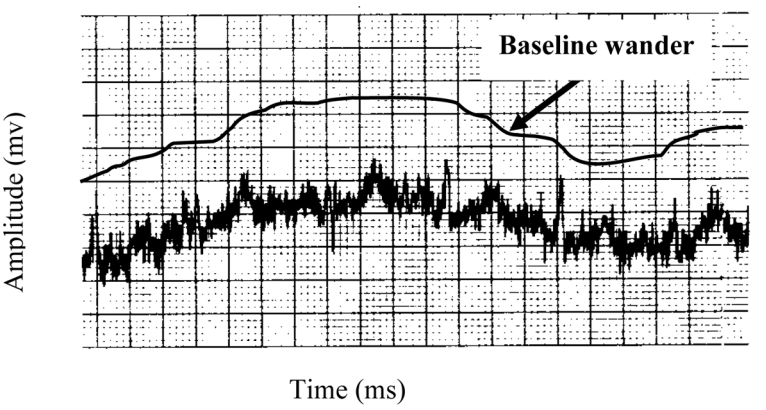


Figure 4. Baseline wander



6.3 EXISTING TECHNIQUES

This part is an overview to the previous works related to different stages of ECG signal processing in order to classify ECG signals to heart disorders.

6.3.1 Noise Removal

There have been several studies on ECG signal noise removal. Zhang and Sui (2010) proposed a method based on morphological filtering and wavelets to eliminate the noise in ECG signals and increase the diagnosis efficiency. In their method, the morphological filter is used to remove the baseline interference signal, and the wavelet transform is applied to remove high frequency interference. Another study was the novel noise-filtering algorithm by Chang (2010), which is based on the ensemble empirical mode decomposition (EEMD) to remove artefacts in ECG traces. Three noise patterns with different characteristics - 50 Hz, Surface electromyograms (EMG) and base line wander- were embedded into simulated and real ECG signals. The mean squared error (MSE) between clean and filtered ECGs was used as the filtering performance index. Results showed that high noise reduction was achieving by the EEMD based filter, especially on arrhythmia ECGs.

In another study, Ling et al. (2008) proposed some fuzzy rules for formulating and integrating different multiwavelets with pre- and post-filters to incorporate expert knowledge at different noise levels. Jesmin et al. (2011) proposed a new ECG denoising method based on the empirical mode

decomposition (EMD). Their method is able to enhance the ECG signal upon removing the noise with minimum signal distortion. Raj and Venkateswarlu (2011) proposed a new denoising method, which uses Undecimated Wavelet Transform to decompose the raw ECG signal. They performed the shrinkage operation to eliminate the noise from the noisy signal. Yan et al. (2002) developed a modified morphological filtering (MMF) technique for signal conditioning in order to accomplish baseline correction and noise suppression with minimum signal distortion.

Poungponsri and Yu (2009) used the wavelet neural network (WNN) for ECG signal modelling and noise reduction. WNN combines the multiresolution nature of wavelets and the adaptive learning ability of artificial neural networks (ANN), and is trained by a hybrid algorithm that includes the Adaptive Diversity Learning Particle Swarm Optimization (ADLPSO) and the gradient descent optimization. Sotos et al. (2007), presented a noise cancellation system suitable for different biomedical signals based on a multilayer ANN. The proposed method consists of a simple structure similar to the Multiple ADAptive LINear Element (MADALINE) neural network. In another study of Sotos et al. (2007), they worked on removing the baseline drift using ANN. The results obtained showed that the ANN-based approach performs better, with respect to baseline drift reduction and signal distortion at the filter output, than traditional methods.

Chawla (2009), Independent Component Analysis (ICA) used for analysis to remove the artefacts and noise from ECG recordings. The reconstructed ECG signals were compared with the original ECG signal and the results showed that there was a significant improvement in signal quality in terms of signal-to-noise ratio (SNR). Zhang and Sui (2010) proposed a method based on morphological filtering and wavelets to eliminate the noise in ECG signals and increase the diagnosis efficiency. Although there are many new methods for noise removal, many systems for ECG signals still use the band-pass filter because of simplicity in implementation and less number of required coefficients (Losada, 2004; Orfanidis, 1996; Lian and Hoo, 2006; Engin, 2004; Minami et al., 1999; Lin et al., 2006; Naghsh-Nilchi and Kadkhodamohammadi, 2008).

Previous research has indicated that the ANN is effective for denoising signals. Therefore, in this project an automatic method for calculating the cutoff frequency is investigated, since finding the correct cutoff frequency is an important issue for noise removal.

6.3.2 ANN Systems for ECG Classification

Classification: enormous volume of data cause that the study becomes complex and difficult. Usually, inaccurate analysis is a high risk. Therefore, computer-based diagnosis and classification of diseases will be very effective (Moein et al., 2008). One of the difficulties in designing an automated ECG analysis system is the large variation in the morphologies of ECG waveforms. The reason is that not only ECG signals are different from patient to patient, but also they are very different within the same patient in the way that they are dissimilar in various beats. Therefore, it is difficult for an ECG beat classifier to be generalised for classification of various patient's ECG waveform, after performing well on the training data (Gacek and Pedrycz, 2012).

Several methods have been proposed for the automatic classification of ECG signals in recent years. Zumray and Tamer (2011) presented a hybrid neural network for ECG classification, where the authors used two feature extraction methods, based on Fourier and wavelet, for classification. The study of Senhadji (1995) introduced the discrete wavelet transform as the feature extractor and linear discriminates as the classifier. Mehmet (2004) studied the application of fuzzy-hybrid neural networks for ECG beat classification. The authors in (Hu. et al., 1997) proposed a local and global classifier, and combined them as the mixture of intelligent methods. Yeap et al. (1990) used a feedforward neural network for classifying the signals, where five features including the QRS width and offset, amplitude of R segment, the T segment slope and the R–R interval duration, were derived, as pertinent features describing the ECG signals.

A wavelet neural network for bundle branch block classification was presented in (Ceylan and Yuksel, 2011), while a wavelet feature extraction combined with fuzzy neural network classification for Premature Ventricular Contraction (PVC) beat classification was introduced in (Shyu et al., 2004). Elif (2010) used a recurrent neural network trained with the Levenberg-Marquardt algorithm for classification of the ECG signals, whereas the authors in (Hosseini and Reynolds, 2001) used a multilayer perceptron (MLP) neural network classifier to achieve an accuracy of up to 88.3% for ECG signal classification. Gholamhosseini et al. (2006) worked on diagnosis of ECG signals using various feedforward neural network architectures. Another research (Ozbay, 2006) presents a comparative study of the classification accuracy of ECG signals using MLP with BP algorithm, and a new fuzzy clustering neural network architecture (FCNN) for early diagnosis.

Inan et al., (2006) combined wavelet-transformed ECG waves with timing information for feature set for classification. The method in (Osowski and Linh, 2001) is based on a hybrid fuzzy neural network that consists of a fuzzy self-organising sub-network connected in cascade with a MLP. The authors proposed to use high-order statistics as input features for feeding their classifier. Ebrahim (2010) used different supervised classifiers for recognition of the PVC from the normal beats and other heart diseases. Ceylan et al. (2009) proposed a diagnostic system consists of combined FCNN algorithm for classification of ECG arrhythmias using type-2 fuzzy c-means clustering (T2FCM) algorithm to improve the performance of neural network.

Literature review shows that previous studies are involved with many mathematical equations and there is a complex set of operations for classification in each of them. It is evident that there is a need for a simpler efficient system for classification of ECG signals for effective and robust heart disorder recognition. In this work, a technique based on the kinetic theory of gas molecules is explored to achieve this objective.

6.3.3 ECG Feature Extraction

This section provides an overview on various techniques and transformations used to extract the features from ECG signals. Zhao et al. (2005) proposed a feature extraction method using wavelet transform and support vector machines. They presented a new approach to in feature extraction for reliable heart rhythm recognition. The proposed classification system is comprised of three components including data preprocessing, feature extraction and classification of ECG signals. Two diverse feature extraction methods are applied together to achieve the feature vector of ECG data. The wavelet transform is used to extract the coefficients of the transform as the features of each ECG segment. Concurrently, autoregressive modelling (AR) is also applied to obtain the temporal structures of the ECG waveforms. Finally, the support vector machine (SVM) with a Gaussian kernel is used to classify different ECG heart rhythms. The results of computer simulations provided to determine the performance of the proposed approach achieved an overall accuracy of 99.68%.

A novel approach by Castro et al. (2000) is proposed for feature extraction that worked based on an algorithm for wavelet transform, for feature extraction from ECG signal and recognition of abnormal heartbeats. The coefficients vector for each PQRST is gained by the optimal wavelet function. The

coefficients of each cycle are divided into three segments related to P wave, QRS complex, and T wave. The summation of the values from these segments provided the feature vectors of single cycles.

Mahmoodabadi et al. (2005) studied on ECG feature extraction using Daubechies Wavelet. Their proposed feature extraction system worked based on the multi-resolution wavelet transform. The ECG signals from Modified Lead II (MLII) were chosen for processing. The wavelet system achieved the sensitivity of 99.18% and positive predictivity of 98%.

A mathematical morphology for ECG feature extraction was proposed by Tadejko and Rakowski (2007). The primary focus of their work is to evaluate the classification performance of an automatic classifier of the ECG for the detection of abnormal beats with a new concept in the feature extraction stage. The obtained feature sets were based on ECG morphology and RR-intervals. The configuration adopted a well known Kohonen self-organising map examination of signal features and clustering.

6.3.4 Swarm Intelligence for Optimization

There has been a large body of work in the area of swarm intelligence for optimization and solving different problems. Various models have been introduced in the past. The Artificial Immune Systems (AIS) are adaptive systems inspired by theoretical immunology and observed immune functions, principles and models, applied to complex problem domains (Farmer et al., 1986). Ant Colony Optimization (ACO) is another algorithm, which is based on the indirect communication between the ants by means of chemical pheromone trails, which enables them to find short paths between their nest and food sources. Dorigo et al. (1996) introduced an Ant System (AS), which is an analogy of Ant Colony Optimization. The main characteristics of their model are positive feedback, distributed computation, and the use of a constructive greedy heuristic.

In the research of Dong et al. (2007), the social foraging behaviour of Escherichia coli bacteria has been used to solve optimization problems. They proposed a hybrid approach involving GA and bacterial foraging (BF) algorithms for function optimization problems. Rashedi et al. (2009), proposed a Gravitational Search Algorithm (GSA), which is an optimization algorithm based on the law of gravity and mass interactions. In the introduced algorithm, the searcher agents are a collection of masses which interact

with each other based on the Newtonian gravity and the laws of motion. In the study of Formato (2008) a new deterministic multi-dimensional search metaheuristic based on the metaphor of gravitational kinematics is proposed and called the Central Force Optimization (CFO).

Zne and Chou (2005) proposed a hybrid search algorithm combining the advantages of GA and ACO that can explore the search space and exploit the best solutions. In another research, Buczak and Uhrig (1996) proposed a novel hierarchical fuzzy-genetic information fusion technique. The reasoning takes place by means of fuzzy aggregation functions, capable of combining information by compensatory connectives that better mimic the human reasoning process than union and intersection, employed in traditional set theories.

Pareja et al. (2008), presented a formal search space representation method in combinatorial optimization problems and also designed a tool for the visualisation of the trajectories of heuristic algorithms based on their method. There are also other visualisation tools for the trace of algorithms in search space in a meaningful way. One tool by Wilensky (1999) has been used for the visualisation of the combinatorial optimization problems execution in a simulated environment designed by Sancho et al. (2005). In the study of Pelta et al. (2005), it was shown that visualisation played an important role in understanding the nature of a problem.

Neural networks have been widely used for ECG beat recognition for classification purpose (Inan et al., 2006; Lin et al., 2006; Hu et al., 1997; Hosseini et al., 2001; Gholamhosseini et al., 2006; Fredric and Soowhan, 1996; Engin, 2004; Elif, 2010; Ebrahim, 2010; Ceylan and Yuksel, 2011; Ceylan et al., 2009). However conventional NN suffers from slow convergence to local and global minima and from random settings of initial values of weights, which may cause the neural networks to suffer from poor mappings of inputs to output. So, researchers have started to use hybrid structures and swarm intelligence methods to improve classification.

Overall, a review of the presented algorithms shows that there is no one superior method for optimising all types of problems. Each algorithm is best suited for solving a special form of problems and it is practical to introduce new algorithms with better performance for solving a special form of problems (Yao et al., 1999). The main aim of the method proposed in this work would be is to increase the performance of the ECG signal classification with minimal complexity.

6.4 SUMMARY

This chapter provides some background and exposure to the details of the ECG waveform, some of heart disorders and different types of noises that corrupt the ECG signal. In addition, the basic concepts of ANN and PSO are explained. Various studies related to different stages of ECG signal classification are also summarised in three sections. The first section explains previous studies for ECG signal noise removal. The second and the third sections review the studies related to the application of various ANN based methods for ECG signal classification and the concepts of SI for optimization problems, respectively.

REFERENCES

Asadi, H., Dowling, R., Yan, B., & Mitchell, P. (2014). Machine learning for outcome prediction of acute ischemic stroke post intra-arterial therapy. *PLoS One*, *9*(2), e88225. doi:10.1371/journal.pone.0088225 PMID:24520356

Birkner, M. D., Kalantri, S., & Solao, V. (2007). Creating diagnostic scores using data-adaptive regression: An application to prediction of 30-day mortality among stroke victims in a rural hospital in India. *Therapeutics and Clinical Risk Management*, *3*, 475–484. PMID:18488068

Eberhart, R., & Shi, Y. (2001). Particle swarm optimization: developments, application and resources. *Proceedings of IEEE Congress of Evolutionary Computation*, 81-86. 10.1109/CEC.2001.934374

Ebrahim, Z. A., Khazaee, A., and Ranaee, V. (2010). Classification of the electrocardiogram signals using supervised classifiers and efficient features. *Comp Meth Prog Biomed*, (99), 179-194.

Elif, D. U. (2010). Recurrent neural networks employing Lyapunov exponents for analysis of ECG signals. *Expert Systems with Applications*, (37): 1192–1199.

Engin, M. (2004). ECG beat classification using neuro-fuzzy network. *Pattern Recognition Letters*, *25*(15), 1715–1722. doi:10.1016/j.patrec.2004.06.014

Farmer, J. D., Packard, N. H., & Perelson, A. S. (1986). The immune system, adaptation and machine learning. *Physica*, (D2), 187–204.

Fitzgibbon, E., Berger, R., Tsitlik, J., & Halperin, H. R. (2002). Determination of the noise source in the electrocardiogram during cardiopulmonary resuscitation. *Critical Care Medicine*, *30*(4Supplement), 148–153. doi:10.1097/00003246-200204001-00006 PMID:11953642

Flake, G. (1999). *The Computational Beauty of Nature*. Cambridge, MA: MIT Press.

Formato, R. A. (2008). Central force optimization: a new nature inspired computational framework for multidimensional search and optimization. *Studies in Computational Intelligence*, (129), 221-238.

Fredric, M. H., & Soowhan, H. (1996). Classification of cardial arrhythmias. *IEEE Transactions on Biomedical Engineering*, *43*(4), 425–430. doi:10.1109/10.486263 PMID:8626192

Hardy, A. (1996). On the number of clusters. *Computational Statistics & Data Analysis*, *23*(1), 83–96. doi:10.1016/S0167-9473(96)00022-9

Hassoun, M. H. (1995). *Fundamentals of Artificial Neural Network*. MIT.

Ho, K. C., Speier, W., & El-Saden, S. (2014). Predicting discharge mortality after acute ischemic stroke using balanced data. *AMIA Annu Symp Proc*, 1787–96.

Hope, T. M., Seghier, M. L., Leff, A. P., & Price, C. J. (2013). Predicting outcome and recovery after stroke with lesions extracted from MRI images. *NeuroImage. Clinical*, *2*, 424–433. doi:10.1016/j.nicl.2013.03.005 PMID:24179796

Hopwood, S. J., & Jeans, J. (2009). *An Introduction to the Kinetic Theory of Gases*. Cambridge University Press.

Hu, G. M., Palreddy, S., & Tompkins, W. (1997). Patient adaptable ECG beat classifier using a mixture of experts approach. *IEEE Transactions on Biomedical Engineering*, (44): 891–900. PMID:9282481

Ismail, W., & Shamsuddin, S. M. (2008). *Differential evolution for neural networks training enhancement* (M.Sc. Thesis). University Technology of Malaysia.

Jesmin, K., Sharif, B., & Gregory, M. (2011). Empirical mode decomposition of the ECG signal for noise removal. *Optical Pattern Recognition*, (8055), 504-512.

Karl, G. R., Isis, A. W., Roberto, L., & Hakan, N. (2004). Fetal ECG waveform analysis. Best Practice & Research. *Clinics in Obstetrics and Gynaecology*, *3*(18), 485–514.

Kennedy, J., & Eberhart, R. (1995). A new optimiser using particle swarm theory. *Proceedings of the 6th International Symposium on Micro Machine and Human Science*, 39-43.

Kim, G. H., Yoon, J. E., An, S. H., Cho, H. H., & Kang, K. I. (2004). Neural network model incorporating: A genetic algorithm in estimating construction cost. *Building and Environment*, *39*(11), 1333–1340. doi:10.1016/j.buildenv.2004.03.009

Leonard, B. L. (2004). *The Kinetic Theory of Gases*. Dover Phoenix Editions.

Lian, Y., & Hoo, P. C. (2006). Digital elliptic filter application for noise reduction in ECG signal. *WSEAS Trans Electron*, *3*(1), 65–70.

Lian, Z., Zhu, F., Guan, Z., & Shao, X. (2008). The analysis of particle swarm optimization algorithm's convergence. *Proceedings of the 7th World Congress on Intelligent Control and Automation*, 623-628.

Lotfi, A. Z. (1969). Biological application of the theory of fuzzy sets and system. *Biocybernetics of the Central Nervous System*, 199-212.

Love, A., Arnold, C. W., & El-Saden, S. (2013). Unifying acute stroke treatment guidelines for a bayesian belief network. *Studies in Health Technology and Informatics*, *192*, 1012. PMID:23920786

Mahmoodabadi, S. Z., Ahmadian, A., & Abolhasani, M. D. (2005). ECG feature extraction using daubechies wavelets. *Proceedings of the fifth IASTED International conference on Visualization, Imaging and Image Processing*, 343-348.

Mehmet, E. (2004). ECG beat classification using neuro-fuzzy network. *Patt Recog Lett*, *25*(15), 1715–1722. doi:10.1016/j.patrec.2004.06.014

Minami, K., Nakajima, H., & Toyoshima, T. (1999). Real-time discrimination of ventricular tachyarrhythmia with Fourier transform neural network. *IEEE Transactions on Biomedical Engineering*, *46*(2), 179–185. doi:10.1109/10.740880 PMID:9932339

Moein, S., Monadjemi, S. A., & Moallem, P. A. (2008). Novel fuzzy-neural based medical diagnosis system. *Int J. Biological and Life Sciences*, *4*(3), 146–150.

Murali, N. V. (2004). *Design of FIR filters with arbitrary responses. Seminar report*. Bombay: Electrical System Group.

Murff, H. J., FitzHenry, F., & Matheny, M. E. (2011). Automated identification of postoperative complications within an electronic medical record using natural language processing. *Journal of the American Medical Association*, *306*, 848–855. PMID:21862746

Ozbay, Y., Ceylan, R., & Karlik, B. (2006). A fuzzy clustering neural network architecture for classification of ECG arrhythmias. *Computers in Biology and Medicine*, *36*(4), 376–388. doi:10.1016/j.compbiomed.2005.01.006 PMID:15878480

Pareja-Tobes, P., Pelta, D., Sancho-Royo, A., & Verdegay, J. L. (2008). Search spaces representation in optimization problems. *Expert Systems with Applications*, *34*(4), 2891–2895. doi:10.1016/j.eswa.2007.05.046

Paul, J. S., Reddy, M. R. S., & Kumar, V. J. (1997). Automatic detection of PVC's using autoregressive models. *19th Ann Intl Conf IEEE EMBS*, 68-71.

Physionet. (2012). *Physiologic signal archives for biomedical research, A database for heart signals*. Retrieved February 20, 2010 from www.physionet.org/physiobank

Raj, V. N. P., & Venkateswarlu, T. (2011). ECG signal denoising using undecimated wavelet transform. *Electronics Computer Technology, 3rd International Conference on*, 94-98 10.1109/ICECTECH.2011.5941808

Roger, A., & Hans-Jakob, S. (2009). Meet the challenge of high-pass filter and ST-segment requirements with a DC-coupled digital electrocardiogram amplifier. *Journal of Electrocardiology*, *46*(6), 574–579. PMID:19700169

Ye, H., Shen, H., & Dong, Y. (2017). *Using Evidence-Based medicine through Advanced Data Analytics to work toward a National Standard for Hospital-based acute ischemic Stroke treatment*. Mainland China.

Zhang, Q., Xie, Y., & Ye, P. (2013). Acute ischaemic stroke prediction from physiological time series patterns. *The Australasian Medical Journal*, *6*(5), 280–286. doi:10.4066/AMJ.2013.1650 PMID:23745149

Chapter 7
Methodology

ABSTRACT

This chapter uses intelligent methods based on swarm intelligence and artificial neural network to detect heart disorders based on electrocardiogram signals. This chapter has introduced the methodology undertaken in the denoising, feature extraction, and classification of ECG signals to four heart disorders including the normal heartbeat. It also presents denoising using intelligent methods.

7.1 INTRODUCTION

In this research, four heart disorders including bundle branch block, supraventricular tachycardia, anterior myocardial infarction (anterior MI) and inferior myocardial infarction (inferior MI), as well as normal ECG, will be classified using proposed intelligent method.

This chapter discusses the methodology to be employed in this research including the proposed method for classification as well as some related theory, with the implementation details given in the next chapter.

7.2 DATA COLLECTION

The biomedical research community uses the Physiobank archive of ECG signals (Physionet, 2012) for testing different studies. Since the collected signals are to be classified in the final stage of this project, having the correct

DOI: 10.4018/978-1-5225-5580-3.ch007

classification of signals is necessary. Therefore, by consultation with heart specialist, 100 ECG signals were selected from three databases of Physiobank: the MIT-BIH Supraventricular Arrhythmia Database, the MIT-BIH Normal Sinus Rhythm Database, and the MIT-BIH Arrhythmia Database. The sampling frequency of the first two databases was 250 Hz, while the signals of the third database were sampled at 360 Hz.

The MIT-BIH Arrhythmia Database came into existence in 1980 for use in evaluating of detectors of arrhythmias, and it has become publicly available It was the first generally available collection of standard test signals and has become a worldwide standard Arrhythmia Database.

The Supraventricular Arrhythmia database includes the examples of supraventricular arrhythmias in the MIT-BIH Arrhythmia Database.

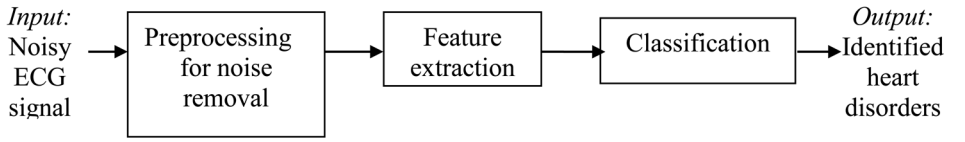
The normal Sinus Rhythm Database includes subjects referred to the Arrhythmia Laboratory at Boston's Beth Israel Hospital that were found to have had no significant arrhythmias; they include 5 men, aged 26 to 45, and 13 women, aged 20 to 50.

7.3 METHODOLOGY STAGES

A database of 100 lead II ECG signals were obtained from the Physiobank database. Figure 1 shows the simplified overall algorithm for the classification.

Preprocessing: The noise from powerline interferences, as well as motion artefacts from the electrode and skin interface, affect the QRS complex, P and T waves of the ECG signals. In the preprocessing, two intelligent approaches based on self-organising map (SOM) and PSO neural network (PSONN) for finding the cutoff frequency, are proposed and applied with one of the traditional models commonly used by many researchers (Losada, 2004; Orfanidis, 1996; Lian and Hoo, 2006; Engin, 2004; Minami et al., 1999; Lin et al., 2006; Naghsh-Nilchi and Kadkhodamohammadi, 2008). In this project, Finite Impulse Response (FIR) filter FIR is preferred to the infinite impulse

Figure 1. Flowchart of overall methodology for ECG classification



response (IIR) filter because IIR filters are more susceptible to problems of finite-length arithmetic. The baseline wander is removed using a median filter as recommended in (Poungponsri and Yu, 2009).

Feature extraction: Feature extraction is the process of reduction of the attributes of data. In feature extraction, the focus is on the similarity of data. After finding the features, it must be decided which are to be consider further as not all features are used. For ECG signal feature extraction, in consultation with heart specialists, it was determined that 6 features are effective for detecting the 4 types of heart disorders (Bundle branch block, supraventricular tachycardia, anterior myocardial infarction (anterior MI), and inferior myocardial infarction (inferior MI)). The 6 features are QRS duration, PR duration, QT duration, heart rate, RR duration and normal ST duration (Positive/Negative T), that will be explained in detail in chapter 9.

Classification: enormous volume of data cause that the study becomes complex and time consuming. Usually, the possibility of missing analysis is high. Therefore, computer-based diagnosis and classification of diseases will be very helpful (Moein et al., 2008). One of the difficulties in designing an automated ECG analysis system is the large variation in the morphologies of ECG waveforms. The reason is that not only ECG signals are different from patient to patient, but also they are very different within the same patient in the way that they are dissimilar in various beats. This is main reason that the beat classifier, performing well on the training data, generalises poorly when presented with different patients' ECG waveforms (Gacek and Pedrycz, 2012).

In the present research, an automated approach based on kinetic energy of gas molecules is proposed. The technique called Kinetic Gas molecule Optimization (KGMO) is then used to train the neural network for classification of ECG signals. Details on feature extraction and classification are described in chapter 5. In the next sections, two proposed approach as for noise removal are described and the test results of the performance are presented.

7.3.1 Noise Removal

In noise removal, two intelligent methods based on the application of SOM and PSONN for ECG signal noise removal are investigated. The performance of the proposed methods is compared with conventionally FIR filtered signals for removing the low amplitude noise. Because of its simplicity in implementation, many other researches (Losada, 2004; Orfanidis, 1996; Lian and Hoo, 2006; Engin, 2004; Minami et al., 1999; Lin et al., 2006; Naghsh-

Nilchi and Kadkhodamohammadi, 2008) have used FIR for noise removal. Also, the median filter is used for removing the baseline wander removal by filtering the high amplitude noise. In the next sections, first there is an explanation on concepts of the FIR filter and median filters and then the proposed intelligent approach for ECG signal noise removal is presented.

7.3.1.1 FIR Filter

Among digital filters, FIR is very attractive for filter designing because of many reasons. Some of the reasons are (Chavan, 2005) simple robustness, easy attaining the linear phase and better support of vendors from FIR hardware.

However, there are some drawbacks in FIR filter. For example, they are computationally expensive to be implemented. The long transient response is another drawback. Usually when the computational power is at a premium, the IIR is used. However, in many cases, the use of multistage/ multirate techniques can yield FIR implementations that can compete (and even surpass) IIR implementations while retaining the nice characteristics of FIR filters such as linear-phase, stability, and robustness to quantization effects (Losada, 2004).

Filter designing starts with the requirements of the desirable FIR. Different methods are used for filter designing depending on the filter specifications and implementation. Choosing the right method of FIR filter designing is very important, as each approach has advantages and disadvantages. One choice for FIR filtering is optimal filtering the given number of impulse response coefficients (Murali, 2004). Another choice is window method that is commonly used method for designing filters because of the simplicity and efficiency. Therefore, the windowing method is used for FIR filtering in this study.

7.3.1.1.1 Windowing

Using the windowing process a small subset of a larger dataset is taken for analysis. In the rectangular window, the dataset is truncated before and after the window, while the content of the window is not modified (Losada, 2004). Applying a window to a dataset will present the spectral properties of that dataset. Consider the system $H(z)$, with input $X(z)$ and output $Y(z)$. We model this as:

$$Y(z) = X(z)H(z) \tag{1}$$

Having a window with transfer function $W(z)$, mathematically the window can be applied to the signal, X(z) as such:

$$\hat{X}(z) = X(z)W(z) \tag{2}$$

Then, the windowed signal can be passed into our system, $H(z)$ as usual:

$$\hat{Y}(z) = \hat{X}(z)H(z) \tag{3}$$

The implementation of the FIR filter requires three parameters: a cutoff frequency (ω_c), the filter order (N), and the window type (W). The filter order primarily determines the width of the transition band. Higher orders give sharper cutoff in the frequency response (Losada, 2004). The final objective of defining filter specifications is to find the desired normalised frequencies (ω_c, ω_{c1}, ω_{c2}), transition width and stopband attenuation. Therefore, the desired sharpness will determine the filtering order. A Hamming window of size $N+1$ is used in the FIR filter, as given by Equation (4).

$$W_n = 0.54 - 0.46\cos(2\pi\frac{n}{N}) \text{ for } 0 \le n \le N \tag{4}$$

7.3.1.2 Median Filter

In a set of numbers, the median is the number for which larger and smaller values are equally probable. To calculate the median for a set of numbers, all the numbers are sorted in any order and then the central value or the mean between the two central values in the even-sized set is calculated.

A moving median is a sliding-window spatial filter that takes a set of points, and given a span for the filter, takes a subset of those points, and returns the median for the subset. It is a discrete time process in which a $2D+1$ points wide window is stepped across an input signal, where D is the sampling frequency. For each window, the output value of the filter is the median of the all ranked points according to their values. Consider a real, discrete-time sequence $\{a(n)\}$ where a is a M-level signal. The output of the median filter $y(n)$ is given by $y(n) = median[a(n-D),...,a(n),...,a(n+D)]$. The median filer for three points would be $y_n = \{median(x(n-1), x(n), x(n+1))\}$.

In comparison with mean, the median is, in a sense, a more robust "average" than the mean, as it is not affected by extreme values. In the next section, the proposed approaches for noise removal are presented.

7.3.1.3 Finding the Cutoff Frequency (ω_c)

Finding the cutoff frequency as the parameter for FIR filtering is described below by two proposed methods based on SOM and PSONN.

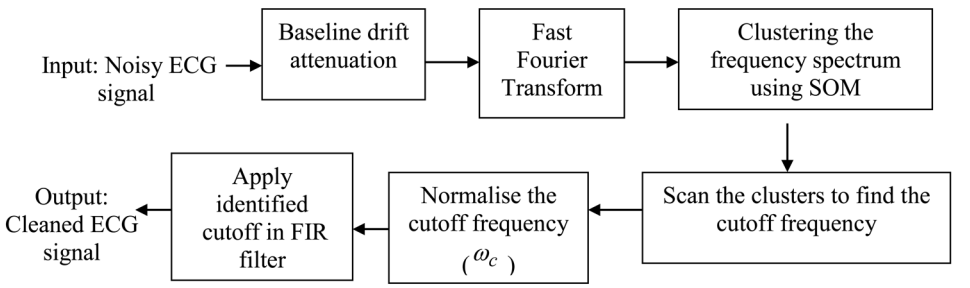
7.3.1.3.1 SOM for Noise Removal

Previous researches have indicated that the artificial neural network (ANN) presents effective approaches for ECG signal classification (Poungponsri and Yu, 2009; Sotos et al., 2007; Sotos et al., 2007). In this section, we present a method using SOM, an unsupervised ANN, to automatically calculate the cutoff frequency. The proposed scheme is discussed in more detail in the next section, followed by a discussion on the results. The proposed method for noise removal is based on clustering the frequency spectrum using SOM. Figure 2 shows the flow of the proposed method.

1. Baseline Drift Attenuation

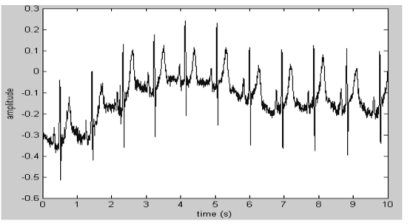
Analysis of typical ECG signals indicates that baseline wander is the prominent phenomenon and its distribution and magnitude varies between different signals. An example is given in Figure 3. The original ECG signal is shown in Figure 3a, where the cycles of the PQRST waveform appear to be attenuated by a sinusoidal waveform. The first step in noise reduction is to attenuate the baseline drift. In previous researches (Roger and Hans,

Figure 2. Flowchart of the proposed noise removal method

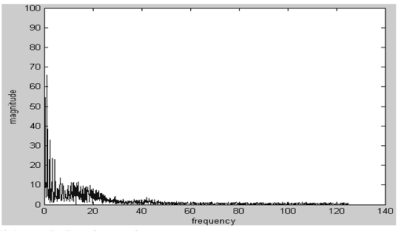


Methodology

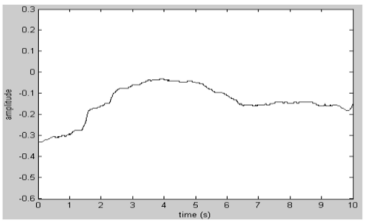
Figure 3. ECG signal drift attenuation and FIR filtering



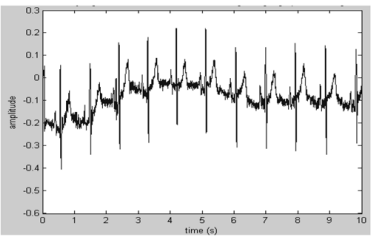
(a) Original ECG signal



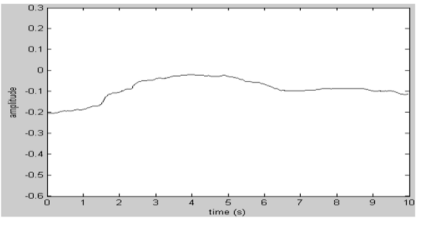
(b) ECG signal Frequency spectrum



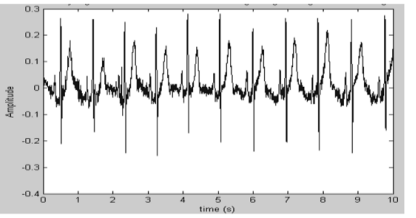
(c) Baseline wander of original signal



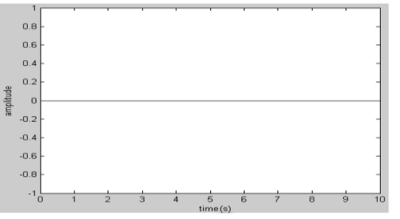
(d) Conventional drift attenuation



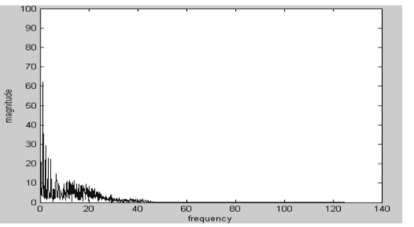
(e) Baseline wander of conventional attenuation



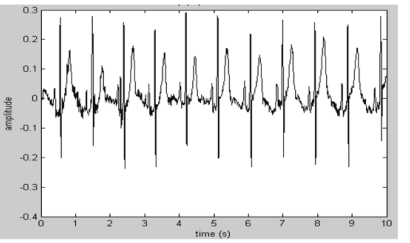
(f) wander removal by median filter



(g) Resulting baseline by median filter



(h) Frequency spectrum after low-pass filtering



(i) Resulting clean ECG signal

2009; Ziarani and Konrad, 2004; Willems and Arnaud, 2002) the high-pass filter with the applied cutoff frequency of 1 Hz (Ziarani and Konrad, 2004; Willems and Arnaud, 2002) is used to attenuate the drift of baseline. Figure 3d shows the ECG signal after low frequency noise reduction with the FIR high-pass filter. Since the moving median filter is more efficient for baseline drift attenuation and in some cases, the high-pass filter causes ST segment distortion (Roger and Hans, 2009), the moving median filter is applied in this work to accomplish this.

Figure 3c shows the extracted baseline. The resulting effect of baseline drift removal is given in Figure 3e, where the baseline wander is eliminated (i.e., virtual straight line at 0 amplitude). Figure 3f is the ECG signal after baseline drift attenuation by median filter with size $2D+1$, and the resulting baseline is shown in Figure 3g.

2. Application of FIR Filtering

After attenuation of the baseline drift, the high amplitude noise will be removed. In order to obtain the spectrum of the ECG signal, it is transformed to the frequency domain using the Fast Fourier Transform (FFT) defined as follows:

$$X_k = \sum^{N-1} x_n e^{-i2\pi k \frac{n}{N}} \text{ for } k = 0,\ldots..,N-1 \tag{5}$$

where $X_0,\ldots..,X_{N-1}$ are complex numbers.

The cutoff frequency is normalised as follows:

$$\omega = \frac{\omega_c}{f_s} \tag{6}$$

where ω is the normalised cutoff frequency and f_s is the sampling frequency.

Figure 3b shows the frequency spectrum up to half of the sampling frequency, which is 360 Hz for the selected ECG signal. As mentioned previously, this study focuses on an intelligent approach for finding this cutoff using SOM neural network. Figure 3h shows the resulting frequency spectrum by the proposed approach and the final ECG signal after FIR lowpass filtering is shown in Figure 3i.

3. Clustering of ECG Frequency Spectrum Using SOM

Kohonen Self Organising Feature Maps, or SOMs (Hassoun, 1995) provides an approach for reducing the dimension of data from multidimensional in much lower dimensional space; usually one or two dimensions. SOM is an unsupervised learning algorithm for modelling the structure of a sample set of patterns. Commonly, SOM is a two-layer neural network: the first being the input layer and the second the competitive layer. The nodes of the first layer selectively feed input elements into the competitive layer of the network. There is a weight vector, which is assigned to each node in the input and competitive layers. Training cycles start with randomly chosen weights for the nodes in the competitive layer (Hassoun, 1995) and the winner node is determined based on Euclidean distance between the weight vector q_i and the input vector x_i such that:

$$\left\|x_v - w_i\right\| = \min\left\|x_v - q_i\right\| \text{ for } i = 1,2,....N \tag{7}$$

where $\left\| \right\|$ indicates Euclidean distance, and x_v indicates the input vector.

The weight vectors of the winning node and the nodes in the neighbourhood are updated using a weight adaptation function based on the following Kohonen rule:

$$\Delta w_i = \alpha(x_v - q_i^{old}) \text{ for } i \in N_r \tag{8}$$

where α is the learning coefficient, q_i^{old} is the weight vector before updating, and N_r is the collection of all nodes in the neighbourhood of radial distance r from the input layer.

In general, the proposed method clusters the ECG frequency spectrum. The input data consists of typical forms of spectrums for four classes of ECG signals, namely the normal sinus rhythm, supraventricular tachycardia, myocardial infarction and Bundle branch block. The critical parameter is the number of clusters for configuration of the SOM. The classical geometric method is a well-known technique, which consists of plotting the value of the clustering criterion, and assessing the plot by analysing discontinuities in the slope (Hardy, 1996). Sharp steps in the curve determine the boundaries of the clusters. Based on analysis of the classical geometric method on the database, it was deduced that the adequate number of clusters was four for

each spectrum. This is seen in Figure 4, which shows typical shapes for the spectrums of the four classes of ECG signals. In the figure, the thick black lines indicate the boundaries of the range of amplitudes for each cluster. The typical ranges for the clusters of the ECG signal were found to be as given in Table 1. The ranges are gained by looking to the amplitude axis and the gradient difference in Figure 4. For instance, the first row of Table 1 is obtained by looking to the normal sinus spectrum in Figure 4a that shows the gradient has changed in ranges 0-5, 6-20, 21-60 and up to 60 (mv). The ranges may vary a little for spectrums in each class, however, the number of clusters was determined to be four for all spectrums.

Therefore, the SOM is configured with a competitive layer with a 1×4 one-dimensional map. The FFT of the ECG signals is the input vector that will be fed to the competitive layer of the SOM. After training, all samples will be clustered in $c_i = i, (i = 1...4)$. Figure 5 shows three ECG signals based on different shapes and characteristics after baseline drift attenuation and the clustering results of FFT using SOM. Since the low amplitude noise are mostly gathered in the lowest cluster, c_1, the cutoff frequency is the point that $c_1 = 1$ changes to $c_2 = 2$ through right to left scanning of the FFT samples.

It is important to note that there must be stability in the clusters when selecting the cutoff frequency. The stability is determined by defining a left neighbourhood border for each sample. In the right to left scanning of the spectrum, the samples are considered stable in a cluster if:

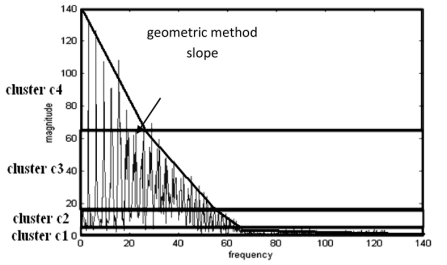
$$\forall s_i \in c_2 : \delta = 10 \Rightarrow n \geq 4 \tag{9}$$

where s_i is the sample of clustered FFT, δ is the left neighbourhood border and $n:\ n \in c_2$ is the number of the consecutive samples at δ. Based on empirical results, the optimum value for δ was found to be 10 samples.

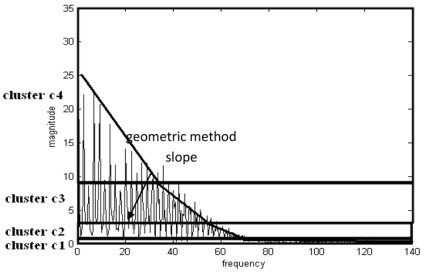
Table 1. Typical ranges of amplitude (mv) for the four classes of ECG

Remark	C4	C3	C2	C1	
See Figure 4a	> 60	21-60	6-20	0-5	**Normal sinus rhythm**
See Figure 4b	> 10	5-10	2-4	0-1	**Supraventricular tachycardia**
See Figure 4c	> 40	16-40	6-15	0-5	**Myocardial infarction**
See Figure 4d	> 70	41-70	6-40	0-5	**Bundle branch block**

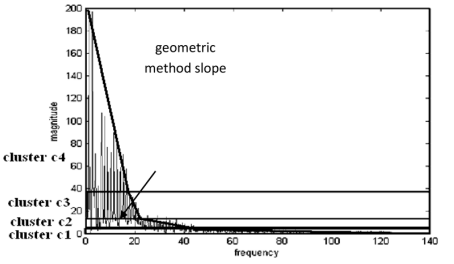
Figure 4. Four typical ranges of amplitude for frequency spectrum in four classes clusters of ECG; the gradient changes between the lines



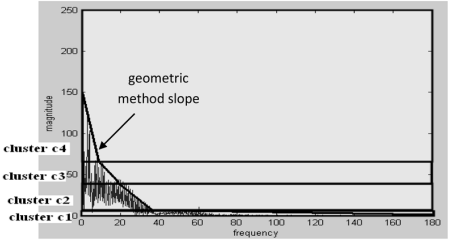
(a) Normal sinus rhythm



(b) Supraventricular tachycardia

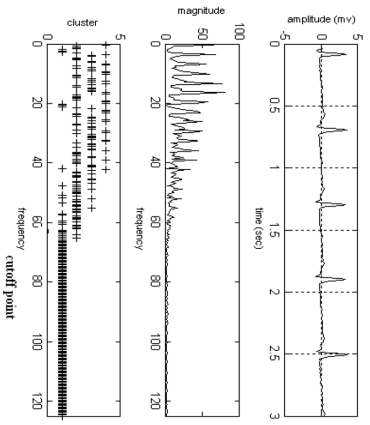


(c) Myocardial infarction

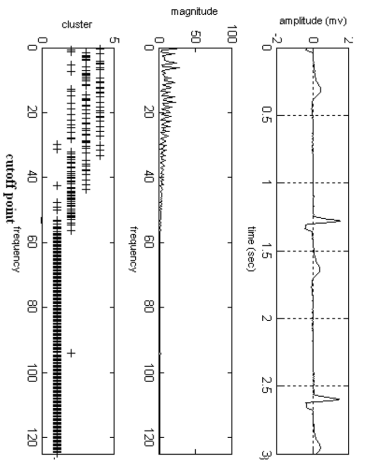


(d) Bundle branch block

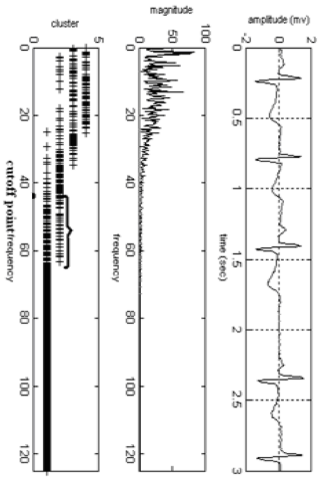
Figure 5. Clustering the frequency spectrum (a. ECG signal after baseline drift removal, b. Signal in frequency domain, c. clustering of FFT)



(a) Example of a normal sinus ECG signal



(b) Example of a supraventricular tachycardia ECG signal



(c) Example of a myocardial infarction ECG signal

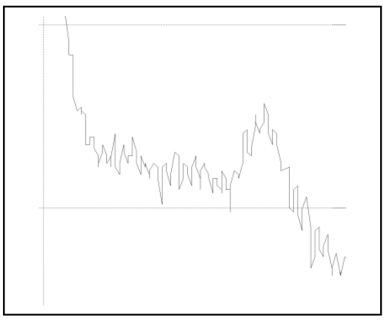
Various tests from 5 to10 samples for δ show that $\delta = 10$ provides better results for finding the cutoff frequency.

Experimental results show that four is the adequate number of samples for determining the cutoff frequency. In other words, the cutoff frequency is where there are more than 4 neighbouring samples in $c_2 = 2$ for $\delta = 10$ through scanning from right to left of the frequency spectrum. For example,

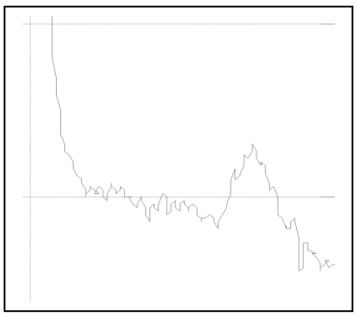
in Figure 5b when scanning from right to left, one encounters a single point first, and since it is less than 4, the mentioned single point is ignored and scanning should be continued until reaching the next set of points. Another example is shown in Figure 5c, which has more single points that are ignored. Again, the scanning process is continued in order to reach to the stable cluster with the mentioned conditions i.e., $\delta = 10 \Rightarrow n \geq 4$. The obtained frequency ω_c is then normalised by equation (6).

Investigation on the output of the first stable point identified still showed influence of noise, as seen in Figure 6b. To overcome this noise (possibly due to equipment/transmission), the next stable point is used. As observed in Figure 6c, this cutoff frequency enables a cleaner signal to be extracted by the FIR filter. The evaluation of the performance with the traditional methods is provided in the next section.

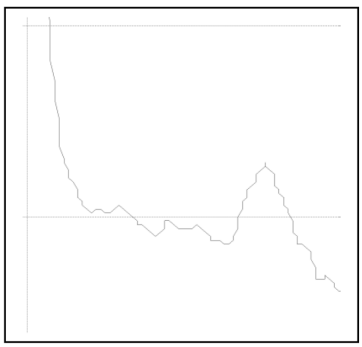
Figure 6. Denoising by SOM based on the stable points



(a) Original ECG signal



(b) Denoising by the first stable point



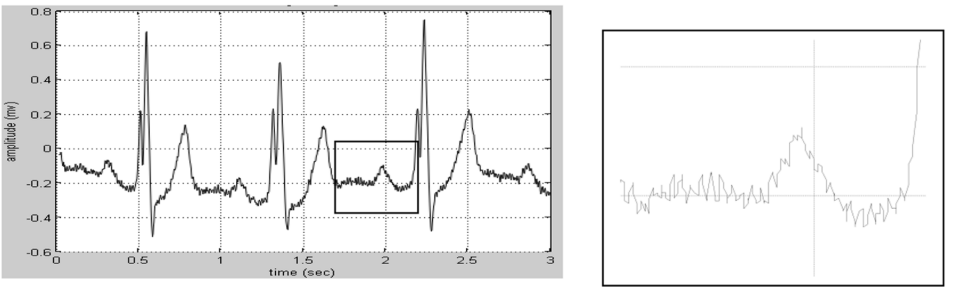
(c) Denoising by the second stable point

4. Test and Results for Noise Removal by SOM Method

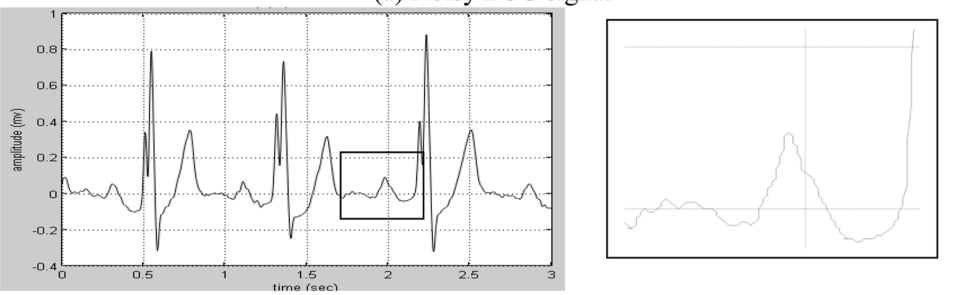
Figures 7 to 10 present 4 examples of ECG signals with different shapes and characteristics from the mentioned databases. The automatically denoised signals are shown in part (b) of the figures. To illustrate the denoising effects more clearly, part of the signal is enlarged on the right of the images. As seen from results, a smooth signal is obtained for each type of disorder. The results of FIR filtering by the calculated cutoff frequency are compared with the results of FIR low-pass filtering based on previous studies (Lian and Hoo, 2006; Engin, 2004; Minami et al., 1999; Lin et al., 2006). In these studies, the cutoff frequency for the ECG low-pass filtering was 100 Hz, where all frequencies above 100 Hz are eliminated to remove power-line interference. For fairness in evaluation, the baseline drift has been attenuated for all the test signals using the moving median filter, instead of just using a 1 Hz high-pass filter as per the norm, since it is has been shown in Figure 3 that the moving median filter is better suited for baseline wander removal. The results of the classical filtering after baseline wander removal are given in part (c) of Figures 7 to 10. Using the proposed method, the calculated cutoff frequency has been found to be equal or better than the classical results for ECG noise removal. Comparing the results, it is seen that in all cases of atrial fibrillation, arrhythmia and supraventricular ECG signals, the proposed intelligently calculated cutoff frequency produced better results than the classical low-pass filter.

In order to properly design the filter, the cutoff frequency for each type of ECG signal should be known and without eliminating signal information, it is impossible to completely remove the noise. In addition, the cutoff frequency is dynamic. Therefore, for various ECG signals, the cutoff frequency would be different due to the biometric characteristics of the ECG signal. In (Sornmo and Laguna, 2005; Konstantions, 2006), the ECG signals varied from person to person. Therefore, it is obvious that the frequency spectrum would be different for each person and the applied cutoff frequency has to vary, instead of being fitted as in the classical low-pass filter where 100 Hz is used as the constant threshold. To better evaluate the proposed scheme, visual identification of the optimum cutoff frequency was undertaken. The cutoff frequencies for the best visual results (up to the point just before peak amplitudes were negatively affected) were identified and the results are shown in part (d) of Figures 7 to 10.

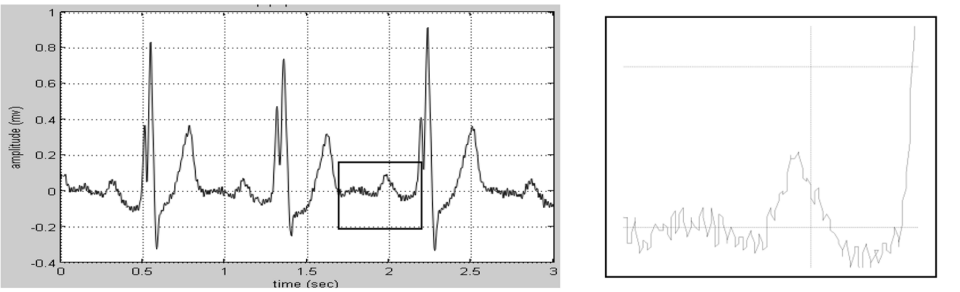
Figure 7. Arrhythmia ECG signals; enlarged portions of the signals are given on the right



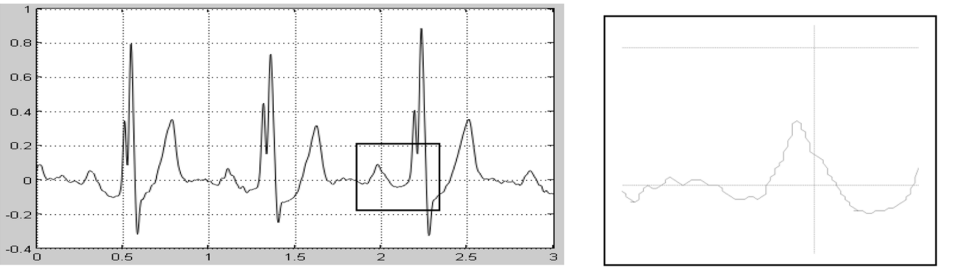
(a) Noisy ECG signal



(b) Denoised ECG signal by the automated calculated cutoff

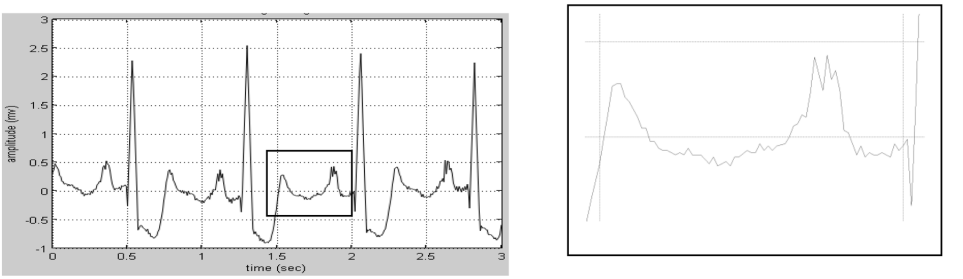


(c) Denoised ECG signal by the classic cutoff (100 Hz)

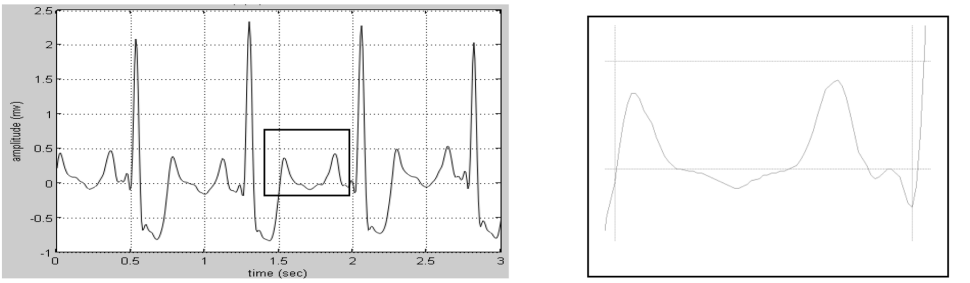


(d) Denoised ECG signal by the visually calculated cutoff

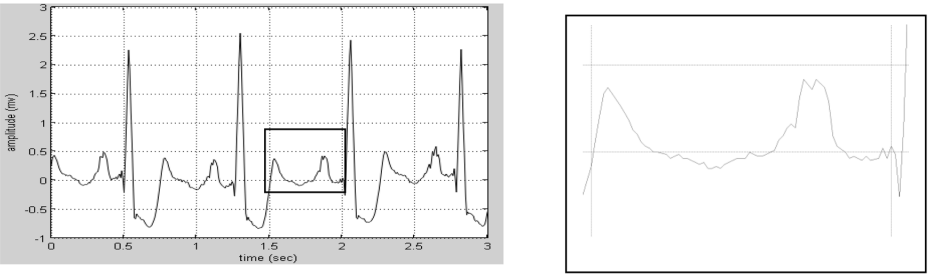
Figure 8. Supraventricular ECG signals; enlarged portions of the signals are given on the right



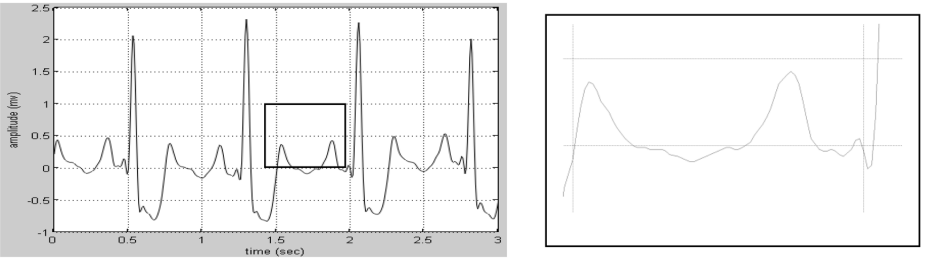
(a) Noisy ECG signal



(b) Denoised ECG signal by the automated calculated cutoff

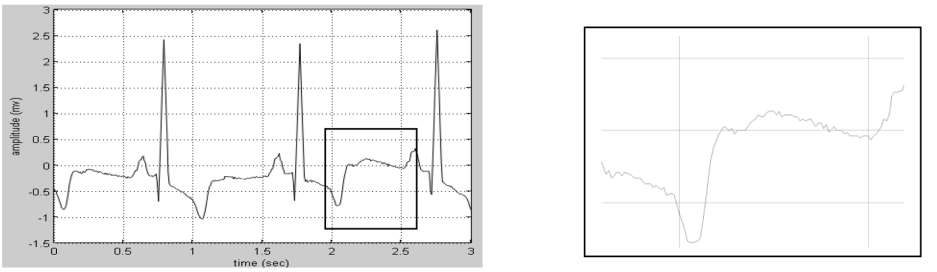


(c) Denoised ECG signal by the classic cutoff (100 Hz)

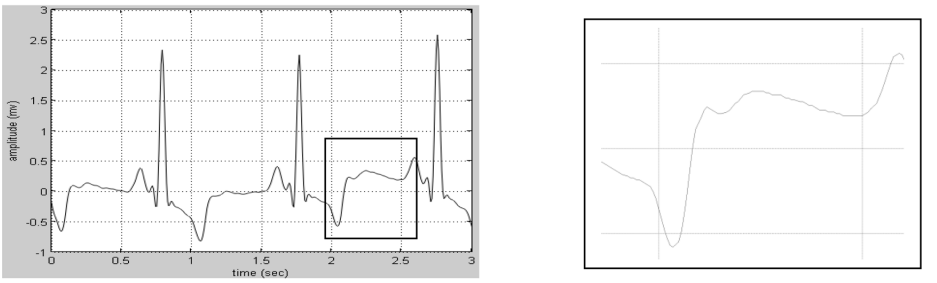


(d) Denoised ECG signal by the visually calculated cutoff

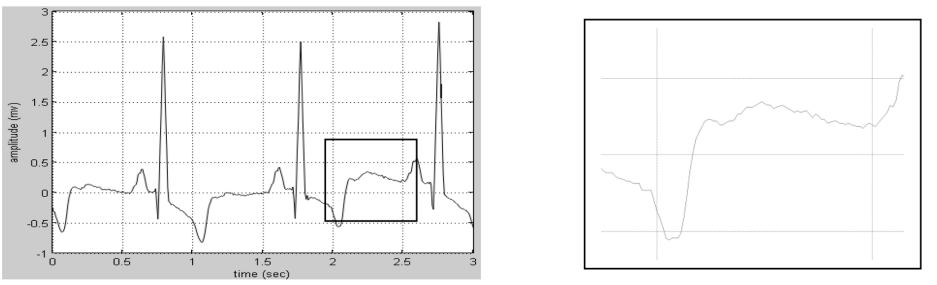
Figure 9. Another supraventricular ECG signals; enlarged portions of the signals are given on the right.



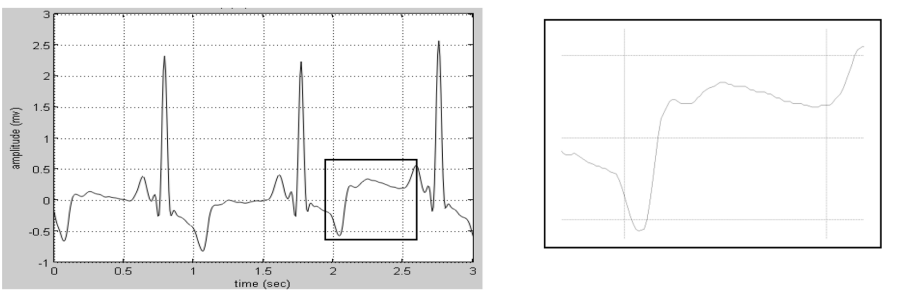
(a) Noisy ECG signal



(b) Denoised ECG signal by the automated calculated cutoff

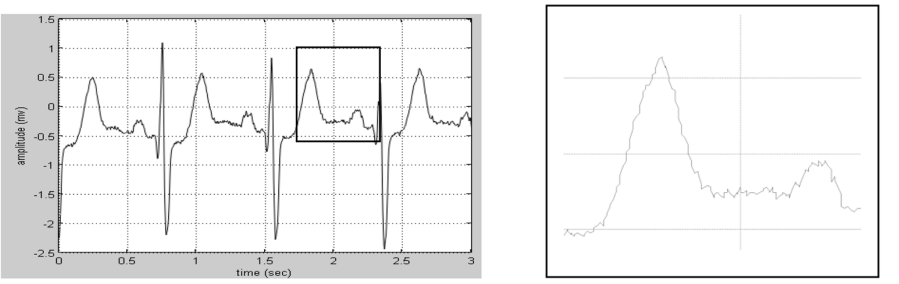


(c) Denoised ECG signal by the classic cutoff (100 Hz)

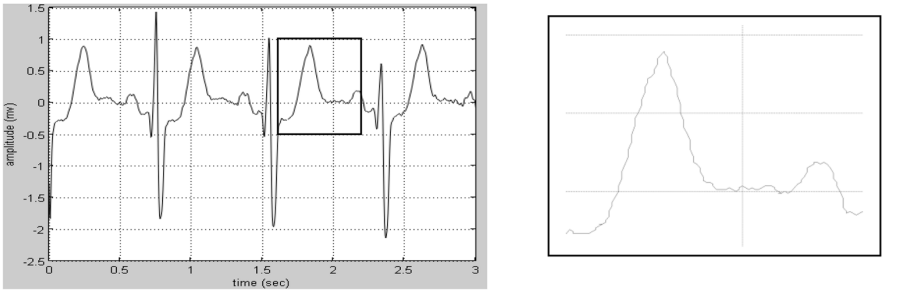


(d) Denoised ECG signal by the visually calculated cutoff

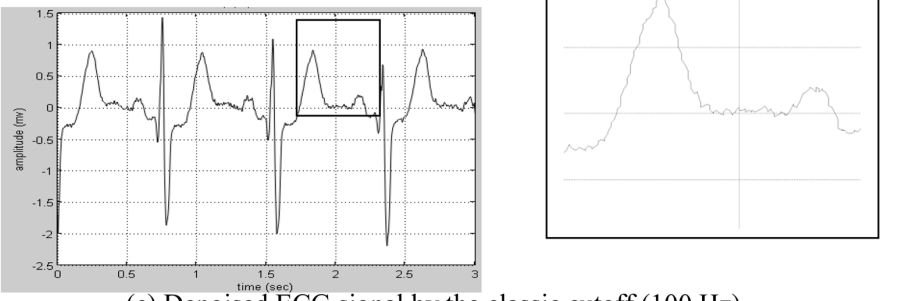
Figure 10. Atrial fibrillation ECG signals; enlarged portions of the signals are given on the right



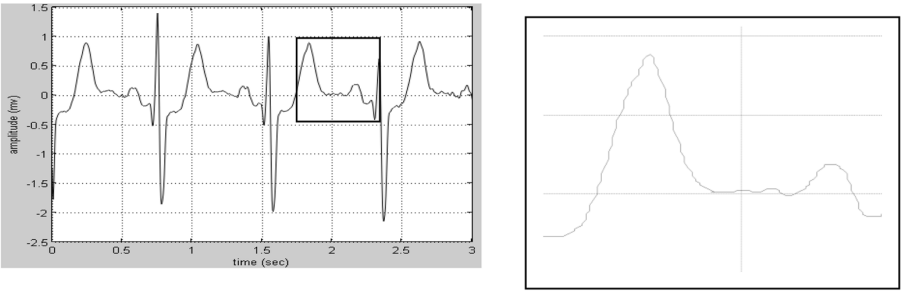
(a) Noisy ECG signal



(b) Denoised ECG signal by the automated calculated cutoff



(c) Denoised ECG signal by the classic cutoff (100 Hz)



(d) Denoised ECG signal by the visually calculated cutoff

The automated ω_c obtained by the SOM for 20 of the ECG signals is given in Table 2. The table shows that even for similar types of heart disorders, there is a significant difference between the cutoff frequencies of the ECGs, further emphasising the difficulty in automatic identification of the optimum cutoff frequency. The visual cutoff frequencies (Visual ω_c) is also given in comparison with the automated cutoff frequencies in the table indicates that, the difference between the automatically determined cutoff frequency and that of the visual examination is small, as given by the mean squared error (MSE). Thus, the viability of the proposed method in correctly determining the cutoff frequency automatically is further supported.

Table 2. Results for comparison between SOM and conventional method for identifying the cutoff frequency (in Hz)

No.	Test signal	Visual ω	Automated ω	Conventional ω
1	Supraventricular (rec-800)	54.1	54.68	100
2	Supraventricular (rec-801)	53.2	53.63	100
3	Supraventricular (rec-802)	64.2	64.84	100
4	Supraventricular (rec-805)	43.6	44.14	100
5	Supraventricular (rec-806)	80.1	80.46	100
6	Supraventricular (rec-807)	38.3	38.67	100
7	Supraventricular (rec-808)	40.1	41.05	100
8	arrhythmia (rec-101)	42.4	43.21	100
9	arrhythmia (rec-111)	26.4	27.35	100
10	arrhythmia (rec-113)	41.6	42.50	100
11	arrhythmia (rec-115)	31.9	32.51	100
12	arrhythmia (rec-116)	32.1	32.90	100
13	atrial fibrillation (rec-iaf1-ivc)	42.3	42.80	100
14	atrial fibrillation (rec-iaf2-ivc)	44.8	45.60	100
15	atrial fibrillation (rec-iaf6-ivc)	23.6	24.20	100
16	atrial fibrillation (rec-iaf13-ivc)	35.1	35.20	100
17	normal (rec-16483)	68.8	69.53	100
18	normal (rec-16420)	82.1	82.81	100
19	normal (rec-16773)	54.6	55.46	100
20	normal (rec-16273)	62.1	62.10	100
MSE			**0.48**	**2958.029**

An intelligent approach based on identifying the cutoff frequency using SOM for ECG low-pass filtering is proposed. The results show that the proposed approach is successful and promising. The results present that the identified cutoff frequency is accurate and better than the conventional method as the SOM-based approach works on each signal individually. Examples of atrial fibrillation, arrhythmia and supraventricular ECG signals show that the low-pass filtering with the automated calculated cutoff frequency generates smoother signal. Also, no significant loss of data is encountered in the denoised signal during the ECG low-pass filtering.

7.3.1.3.2 Particle Swarm Optimization Neural Network (PSONN) for Noise Removal

This section presents the application of PSO in training the ANN to identify the optimal cutoff frequency for removal of the high frequency noise in ECG signals using an FIR filter. Some of the attractive features of the PSO include the ease of implementation and the fact that no gradient information is required. It can be used to solve a wide array of different optimization problems, including most of the problems that can be solved using Genetic Algorithms (GA) (Dong, 2007). Below is a comparison between GA and PSO.

PSO and GA have some similarities. These two optimization algorithms that start the searching process based on the population of random solutions (Eberhart et al., 2001), and for evaluation the solution they use a fitness values. PSO and GA algorithms do not guarantee the success for finding the optimum. However, there are some differences between GA and PSO. For example, unlike GA there is a velocity equation in PSO for updating the solutions. All particles move in the search space with the velocity, which is updated according to its historical behaviour. Therefore, PSO particles tend to move toward the better place in the search area.

Unlike PSO, GA have some operators such as crossover and mutation. Particles update themselves with the internal velocity. PSO particles have memory for keeping the velocity for updating their place in search area. In GA, chromosomes are responsible for sharing information. Therefore, all the chromosomes move toward the optimum solution. PSO has a one way information sharing mechanism which is based on the movement of the global best movement for giving information to others. Table 3 briefly shows the differences of GA and PSO. A more complete table for comparison of GA and PSO is provided in Table A in Appendix A.

Table 3. The differences of GA and PSO

PSO	GA
Particles update themselves with the internal velocity.	Genetic operators are crossover and mutation
Fast convergence	Slower than PSO in converging
Simple operations	Complex operations
Particles move to converge to the best position	No assurance to converge to the best solution

The PSO has been applied to a vast number of problems. This section will briefly describe some of the applications that can be found in the literature (Fukuda et al., 1998; Flake, 1999).

Neural network training was one of the first applications of the PSO. Kennedy and Eberhart (1995) reported that the PSO was successful in training a network to correctly classify the XOR problem, a process involving the minimisation of function in a 13-dimensional search space. They also reported that the PSO could train a neural network to classify Fisher's Iris dataset (Kennedy and Eberhart, 1995), although few details were provided. Salerno (1997) also applied the PSO to the task of training a neural network to learn the XOR problem, reporting significantly better performance than that obtained with a Gradient Descent algorithm. He showed that the PSO was able to train a simple recurrent neural network.

In fact, most PSO applications reported in the literature involve neural network training. Earlier versions of the PSO, before the introduction of the inertia weight or constriction factor, did not have the ability to perform a fine-grained search for the error surface. This led to experiments involving a hybrid between PSO and traditional gradient techniques. Bergh (2001) used the PSO to find a suitable starting position for the Scaled Conjugate Gradient algorithm. Results showed that the hybrid method resulted in significantly better performance on classification problems.

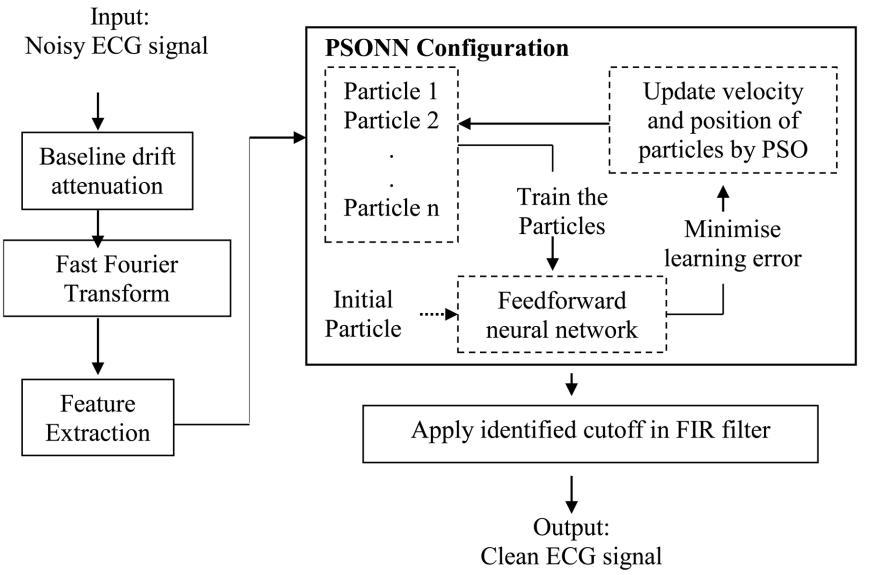
In the proposed approach, PSO is applied for finding the cutoff frequency of the ECG signal for FIR filtering. Figure 11 shows the proposed approach. The main stages are described below.

The collected signals are then transferred to the frequency domain.

1. The Frequency Spectrum Feature Extraction

The inputs of the PSONN are determined by the features of the dataset. Analysis of the ECG signal, taking into account minimal complexity in

Figure 11. Flowchart of the proposed method



computation, revealed that two common statistical measures Root Mean Square (RMS) and standard deviation) would suffice. RMS and standard deviation are commonly used for measuring the dispersion. There is an advantage of using the RMS as it is the sum of squaring of each sample. The squaring makes each term positive, so that the values more than the mean do not cancel values below the mean. Because the RMS is squared, the unit of data is not the same as the unit of variance.

A sample of the value classes in the dataset are given in Table 4. The last column of the table lists the approximate optimum cutoff frequency of each spectrum for FIR low-pass filtering, as determined visually by collaborating signal specialists. This would be used as the target in the evaluation of the intelligent PSONN cutoff frequency determination system.

2. PSONN Configuration

Most of the PSO versions available in literature use a uniform probability distribution to generate random numbers using an inbuilt command such as ‘rand()’ (in C++) available in most of the programming languages.

Table 4. Example of dataset for training the PSONN

No.	FFT RMS	FFT standard deviation	ω_c
1	11.4518	10.8775	33
2	11.4518	10.8775	33
3	11.8526	11.3043	33
4	12.377	11.9541	33
5	12.377	11.9541	33
6	17.5913	16.0341	60
7	17.8998	16.8392	60
8	17.1578	14.995	60
9	18.5355	16.5922	60
10	16.7874	15.0489	60
11	18.0635	16.647	60
12	6.3927	5.3498	100
13	4.7166	3.9738	100
14	9.5551	7.1651	100
15	22.1552	18.3218	80
16	21.3311	16.3213	80
17	20.2209	17.167	80
18	30.1121	20.129	20
19	29.4327	21.6642	20
20	28.1143	19.0133	20

In the classification problems, the number of nodes in each layer is different depending on the type of the problem. Usually the number of nodes in the input layer and output layer depend on the number of attributes and the class attributes, respectively. Based on Kim et al. (2004) there is no specific rule for determining the number of nodes in the hidden layer. There are many suggestions by various researchers that some of them are listed as follows. For example, the number of nodes in the hidden layer should be between the number of nodes in the input layer and the number of nodes in the output layer. Other suggestion is adding the number of nodes in the output layer with 2/3 of number of nodes in the input layer. One suggestion is considering less than twice the number of nodes in the input layer for the number of nodes in the hidden layer.

Based on the theory of Kolmogorov, for n nodes in the input layer and one hidden layer, $2n+1$ nodes in the hidden layer are enough (Shamsuddin, 2004). Charytoniuk and Chen (2000) mentioned that the number of nodes in the hidden layer should be selected based on trial and error approaches.

In this work, the third theory (considering less than twice the number of nodes in the input layer) is used for selecting the number of nodes in the hidden layer of PSONN, as it is more common among researchers who have worked in this area. The PSONN employed to calculate the cutoff frequency is a 3-layer feedforward network. Two nodes are used in the input layer as there are 2 features for each spectrum. Two nodes in the hidden layer and 1 node in the output layer are selected for the 5 classes in the output. The population size used was 25, with the default value of 2 for the maximum velocity divisor, c_1 and c_2. The next section provides the test results achieved for automatic identification of the cutoff frequency by the PSONN and the FIR filter in denoising the ECG signals.

7.3.3.2.3 Test and Results for Noise Removal by PSONN-FIR Method

The proposed system performance using the configured PSONN is tested by measuring various parameters. The FIR is implemented with the cutoff frequency identified by the PSONN and the results are evaluated against the clean and conventionally filtered ECG signals.

Due to the lack of cases in the configured dataset, validity testing of the results was undertaken using a k-folding method with k=3. For each of the k times of NN training, 60% of the samples in the dataset were used for training with the non-overlapping 40% for testing. All the reported results are obtained by averaging the outcomes of three separate tests.

1. Test with Clean Signal

Clean signals would allow for more accurate evaluation of the effectiveness of the proposed method in terms of identification of appropriate cutoff values. As all the signals in the test databases are clinical data that are influenced by noise, and there is no access to actual clean signals, pseudo-clean signals from the ECG generator of the MIT-BIH database (Physionet, 2012) are used. The noisy versions of those pseudo-clean signals are simulated by adding a small value of high frequency components of random noise to mimic the natural noise by the electromygraphic interference, which corrupts real ECG signals.

Figure 12 and Figure 13 show two examples of the tested clean ECG signals. For the ECG signal spectrum in Figure 12, the standard deviation and RMS are 20.54 and 7.12, respectively. For Figure 13, the standard deviation is 68.29 and the RMS is 5.56. The trained PSONN gives the cutoff frequency of 14% and 17% for the first and second ECG signals, respectively. For both figures, part (a) is the pseudo-clean signal, (b) is the signal corrupted by random noise and (c) shows the result of FIR filtering with the cutoff calculated by the proposed method using PSONN. Comparing the pseudo-clean and PSONN filtered ECG signals in both figures show that PSONN can very smoothly filter the high frequency noise and has satisfactory performance for ECG noise removal.

Figure 12. Test results with pseudo-clean ECG signal

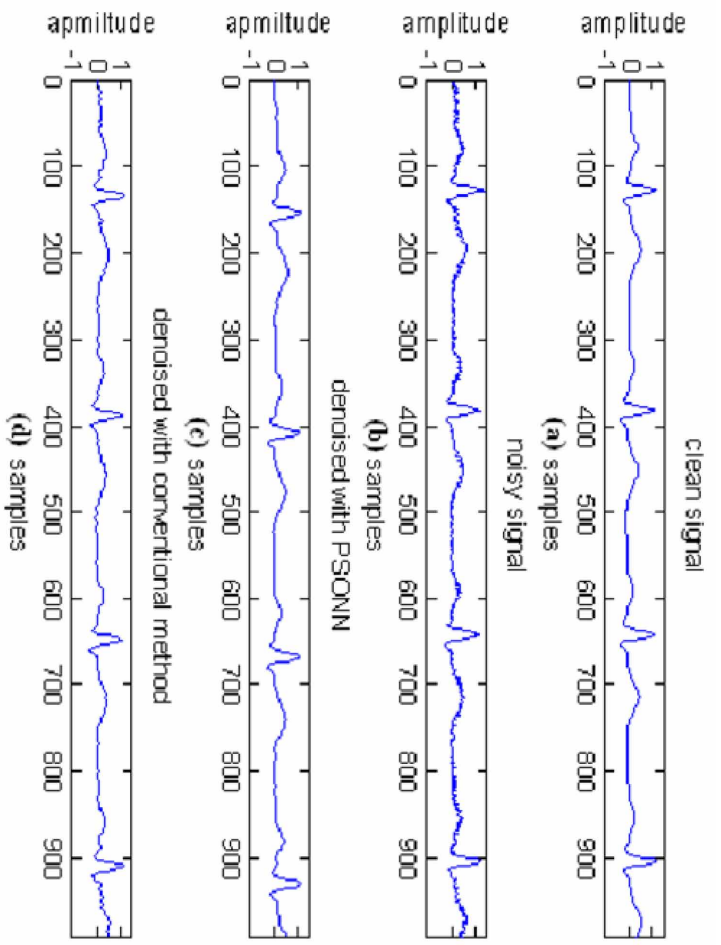
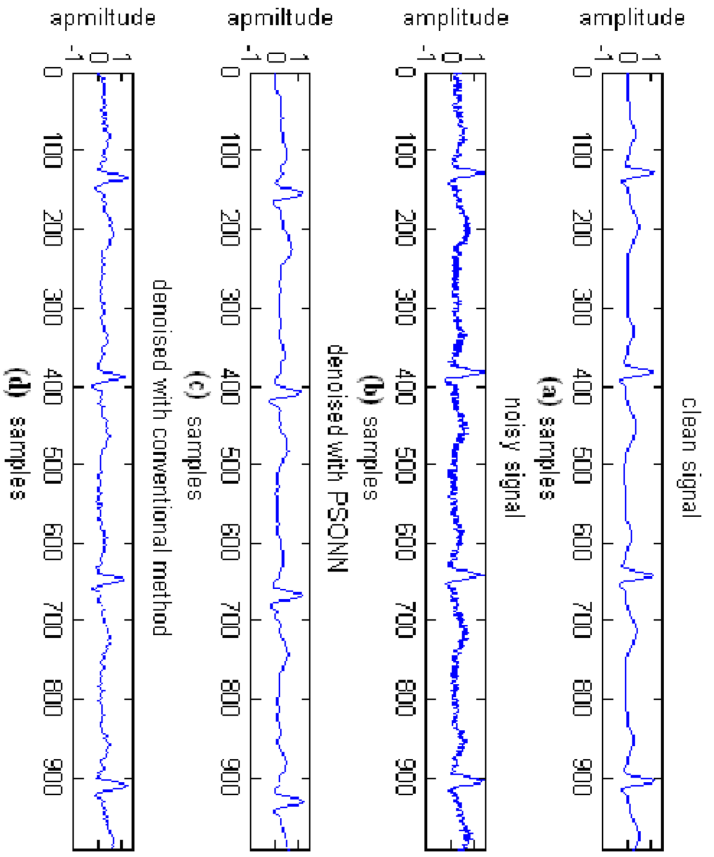


Figure 13. Test results with another pseudo-clean ECG signal



2. MSE Performance

The mean squared error (MSE) is measured for varying numbers of hidden nodes and number of training epochs. To decrease the complexity of the network and increase training speed, less nodes in the hidden layer is preferred (Bergh, 2001). The MSE equation is given by Equation (14).

$$MSE = \frac{1}{N}\sum_{n=0}^{N-1}(y_n - t_n)^2 \tag{14}$$

where y_n is the output of the network, t_n is the desired target, N is the number of test records.

Figure 14 shows the MSE results achieved. It shows that the PSONN has the MSE of 0.16 nodes with 2 and 3 nodes in hidden layer. However, the convergence rate with 1 node in hidden layer is around 200 training cycles, which is faster than PSONN with 2 and 3 nodes in hidden layer. The training will be stopped when the MSE has not changed by at least 1e-009 for 400 epochs.

3. Testing of PSONN against Conventional Filter

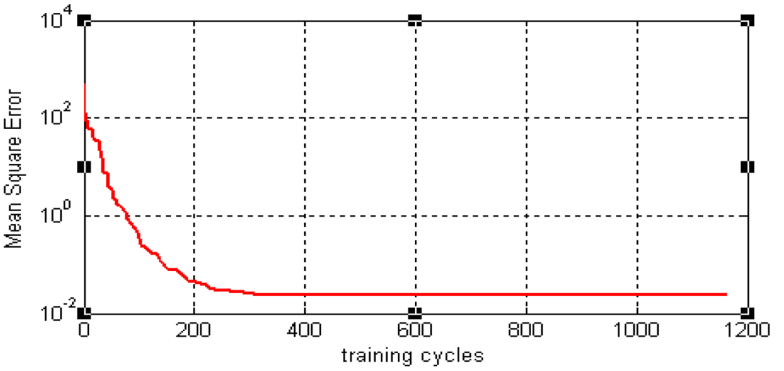
The results of FIR filtering by the calculated cutoff frequency are compared with the results of FIR low-pass filtering based on previous studies (Lian and Hoo, 2006; Engin, 2004; Minami, 1999). In these studies, the cutoff frequency for the ECG low-pass filtering was 100 Hz, where all frequencies above 100 Hz are eliminated to remove high frequency noise. For fairness in evaluation, the baseline drift has been attenuated for all the test signals using the moving median filter, instead of just using a 1 Hz high-pass filter as per the norm, since it has been shown in Figure 3, the moving median filter is better suited for baseline wander removal.

The cutoff frequency is dynamic. Therefore, for various ECG signals, the cutoff frequency would be different due to the biometric characteristics of the ECG signal. In (Sornmo and Laguna, 2005; Konstantinos et al., 2006) the ECG signals varied from person to person. Therefore, it is obvious that the frequency spectrum would be different for each person and the applied cutoff frequency has to vary, instead of being fitted as in the classical low-pass filter where 100 Hz is used as the constant threshold. Figure 12(d) and Figure 13(d) show results of the conventional based filtering of ECG. It presents that compared to the PSONN, the conventional method has low effectiveness in removing the high frequency noise in ECG signals, as little smoothness is observed in the filtered ECG.

This section has shown that PSONN is a reliable and fast solution for automatically identifying the cutoff frequency, that can converge in less than 2000 training epochs. The results show that the PSONN is better able to remove the noise than the conventional method. Also, comparing the clean and PSONN filtered signals presents that the proposed method is promising for use by medical experts who wish to diagnose heart disorders using ECG.

Figure 14. Performance of PSONN with different number of nodes in hidden layer

(a) 1 node in hidden layer, MSE= 0.16



(b) 2 nodes in hidden layer, MSE= 0.16

(c) 3 nodes in hidden layer, MSE= 0.16

7.4 SUMMARY

This chapter has introduced the methodology undertaken in the denoising, feature extraction and classification of ECG signals to four heart disorders including the normal heartbeat. Details on how to remove the high frequency components of noise including the baseline drift attenuation has been explained and then the method has been tested on different signals and evaluated by visual assessment and MSE. Results show that both SOM and PSONN have good performance for noise removal. Details on feature extraction and classification are described in Chapter 8. In the next chapter, KGMO is introduced in detail and then tested on benchmark functions to be evaluated.

REFERENCES

Kirkpatrik, S., Gellat, C.D., & Vecchi, M. P. (1983). *Optimization by simulated annealing*. Academic Press.

Kohavi, R., & Provost, F. (1998). Glossary of Terms. *Machine Learning*, *30*(2/3), 271–274. doi:10.1023/A:1017181826899

Konstantinos, N. P., Dimitrios, H., & Jimmy, K. M. L. (2006). *ECG biometric recognition without fiducial detection, Biometric Consortium Conference*. IEEE Biometrics Symp, Baltimore, MD.

Leonard, B. L. (2004). *The Kinetic Theory of Gases*. Dover Phoenix Editions.

Lian, Y., & Hoo, P. C. (2006). Digital elliptic filter application for noise reduction in ECG signal. *WSEAS Trans Electron*, *3*(1), 65–70.

Lian, Z., Zhu, F., Guan, Z., & Shao, X. (2008). The analysis of particle swarm optimization algorithm's convergence. *Proceedings of the 7th World Congress on Intelligent Control and Automation*, 623-628.

Lin, H., Wensheng, H., Xiaolin, Z., & Chenglin, P. (2006). Recognition of ECG patterns using artificial neural network. *Proceedings of the Sixth IEEE Intl Conf Intell, Sys Design and Appl*, 477-481.

Ling, B. W. K., Ho, C. Y. F., Lam, H. K., Wong, T. P. L., Chan, A. Y. P., & Tam, P. K. S. (2008). Fuzzy rule based multiwavelet ECG signal denoising. *IEEE World Cong Comput Intell*, 1064-1068.

Losada, R. A. (2004). Design finite impulse response digital filters. Part II. *Microwaves & RF*, (43): 70–84.

Lotfi, A. Z. (1969). Biological application of the theory of fuzzy sets and system. *Biocybernetics of the Central Nervous System*, 199-212.

Mahmoodabadi, S. Z., Ahmadian, A., & Abolhasani, M. D. (2005). ECG feature extraction using daubechies wavelets. *Proceedings of the fifth IASTED International conference on Visualization, Imaging and Image Processing*, 343-348.

Mahmoudi, T. M., Forouzideh, N., Lucas, C., & Taghiyareh, F. (2009). Artificial neural network weights optimization based on imperialist competitive Algorithm. *Proceedings of the Seventh International Conference on Computer Science and Information Technologies*, 118-121.

Moein, S., Monadjemi, S. A., & Moallem, P. A. (2008). Novel fuzzy-neural based medical diagnosis system. *Int J. Biological and Life Sciences*, *4*(3), 146–150.

Murali, N. V. (2004). *Design of FIR filters with arbitrary responses. Seminar report*. Bombay: Electrical System Group.

Naghsh-Nilchi, A. R., & Kadkhodamohammadi, A. R. (2008). *Cardiac arrhythmias classification method based on MUSIC morphological descriptors, and neural network. EURASIP J Adv Sig Process*.

Owis, M. I., Youseff, A. B. M., & Kadah, Y. M. (2002). Characterisation of electrocardiogram signals based on blind source separation. *Medical & Biological Engineering & Computing*, *40*(5), 557–564. doi:10.1007/BF02345455 PMID:12452417

Ozbay, Y., Ceylan, R., & Karlik, B. (2006). A fuzzy clustering neural network architecture for classification of ECG arrhythmias. *Computers in Biology and Medicine*, *36*(4), 376–388. doi:10.1016/j.compbiomed.2005.01.006 PMID:15878480

Pareja-Tobes, P., Pelta, D., Sancho-Royo, A., & Verdegay, J. L. (2008). Search spaces representation in optimization problems. *Expert Systems with Applications*, *34*(4), 2891–2895. doi:10.1016/j.eswa.2007.05.046

Paul, J. S., Reddy, M. R. S., & Kumar, V. J. (1997). Automatic detection of PVC's using autoregressive models. *19th Ann Intl Conf IEEE EMBS*, 68-71.

Pelta, D., Krasnogor, N., Bousono-Calzon, C., Verdegay, J. L., Hirst, J., & Burke, E. K. (2005). A fuzzy sets based generalization of contact maps for the overlap of protein structures. *Journal of Fuzzy Sets and Systems*, (152), 103-123.

Physionet. (n.d.). *Physiologic signal archives for biomedical research, A database for heart signals*. Retrieved from www.physionet.org/physiobank

Pomi, A., & Olivera, F. (2006). *BMC Medical Informatics and Decision Making, Context-sensitive auto associative memories as expert systems in medical diagnosis*. BioMed Central.

Poungponsri, S., & Yu, X. H. (2009). Electrocardiogram (ECG) signal modelling and noise reduction using wavelet neural networks. *Proceedings of the IEEE Intl Conf Autom Logistics*, 394-398.

Rashedi, E., Nezamabadi-pour, H., & Saryazdi, S. (2009). GSA: A gravitational search algorithm. *Information Sciences*, *179*(13), 2232–2248. doi:10.1016/j.ins.2009.03.004

Robert, M., & Ian, R. G. (2006). The effects of noise on computerised electrocardiogram measurements. *Journal of Electrocardiology*, *39*(4), 165–173. doi:10.1016/j.jelectrocard.2006.05.025

Salerno, J. (1997). Using the particle swarm optimization technique to train a recurrent neural network. *Proceedings of ninth IEEE international conference on tools with artificial intelligence*, 45-49. 10.1109/TAI.1997.632235

Sancho-Royo, A., Pelta, D., Verdegay, J. L., & Gonzalez, J. R. (2005). Evaluation of co-operative strategies in optimization problems. *Proc. Nisis*. Retrieved from http://www.nisis.de

Senhadji, L., Carrault, G., Bellanger, J. J., & Passariello, G. (1995). Comparing wavelet transforms for recognising cardiac patterns. *IEEE Eng Med Biol*, *14*(2), 167–173. doi:10.1109/51.376755

Shaw, E. (1962). The schooling of fishes. *Scientific American*, *206*(6), 128–138. doi:10.1038cientificamerican0662-128 PMID:14458553

Shyu, L. Y., Wu, Y. H., & Hu, W. C. (2004). Using wavelet transform and fuzzy neural network for VPC detection from the Holter ECG. *IEEE Transactions on Biomedical Engineering*, *51*(7), 1269–1273. doi:10.1109/TBME.2004.824131 PMID:15248543

Sordo, M. (2002). *Introduction to Neural Networks in Healthcare. OpenClinical: Knowledge Management for Medical Care*. Harvard.

Sotos, J. M., Sanchez, C., Mateo, J., Alcaraz, R., Vaya, C., & Rieta, J. J. (2007). Neural networks based approach to remove baseline drift in biomedical signals. *11th Mediterr Conf Med Biomed Eng Comput 2007. IFMBE Proceedings*, *16*(2), 90–93.

Sotos, J. M., Melendez, C. S., Salort, C. V., Abad, R. C., and Ibanez, J. J. R. (2007). A learning based Widrow-Hoff Delta algorithm for noise reduction in biomedical signals, Bio-inspired Modelling of Cognitive Tasks. *Lecturer Notes in Computer Science*, (4527), 377-386.

TomasB. G. (n.d.). Retrieved from http://www.amazon.co.uk/Introduction-12-Lead-ECG-Interpretation-Garcia/dp/0763719617

Neil, E. H. (2002). *Introduction to 12-Lead ECG: The Art of Interpretation*. Jones and Bartlett Publishers.

Tadejko, P., & Rakowski, W. (2007). Mathematical morphology based ECG feature extraction for the purpose of heartbeat classification. *6th International Conference on Computer Information Systems and Industrial Management Applications*, 322-327. 10.1109/CISIM.2007.47

Van, D. B., & Engelbrecht, A. P. (2000). Cooperative learning in neural networks using particle swarm optimisers. *South African Comput*, (26), 84-90.

Venter, G., & Haftka, R. T. (2010). Constrained particle swarm optimization using a bi-objective formulation. *Structural and Multidisciplinary Optimization*, (40), 65-76. Retrieved from http://www.springerlink.com/content/1615-147x/40/1-6/

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, *1*(6), 80–83. doi:10.2307/3001968

Willems, J. L., Arnaud, P., Van Bemmel, J. H., Bourdillon, P. J., Degani, R., Denis, B., ... Zywietz, C. (1987). A reference data base for multi-lead electrocardiographic computer measurement programs. *Journal of the American College of Cardiology*, *10*(6), 1313–1321. doi:10.1016/S0735-1097(87)80136-5 PMID:3680801

Yao, X., Liu, Y., & Lin, G. (1999). Evolutionary programming made faster. *IEEE Transactions on Evolutionary Computation*, (3): 82–102.

Yeap, T. H., Johnson, F., & Rachniowski, M. (1990). ECG beat classification by a neural network. *Proc Ann Intl Conf IEEE EMBS*,1457-1458. 10.1109/IEMBS.1990.691837

Yu, J., Wang, S., & Xi, L. (2008). Evolving artificial neural networks using an improved PSO and DPSO. *Neurocomputing*, *71*(4-6), 1054–1060. doi:10.1016/j.neucom.2007.10.013

Ying, P. C., & Pei, J. (2010). Analysis of particle interaction in particle swarm optimization. *Theoretical Computer Science*, (411): 2101–2115.

Zhan, Z. H., Zhang, J., Li, Y., & Chung, H. S. H. (2009). Adaptive Particle Swarm Optimization. *IEEE Transactions on Systems, Man, and Cybernetics. Part B, Cybernetics*, *39*(6), 1362–1381. doi:10.1109/TSMCB.2009.2015956 PMID:19362911

Ziarani, A. K., & Konrad, A. (2004). A nonlinear adaptive method of elimination of powerline interferences in ECG signals. *IEEE Transactions on Biomedical Engineering*, *49*(6), 540–547. doi:10.1109/TBME.2002.1001968 PMID:12046699

Zumray, D., & Tamer, O. (2001). ECG beat classification by a novel hybrid neural network. *Comp Meth Prog Biomed*, (66), 167-181.

Chapter 8
Kinetic Gas Molecule Optimization (KGMO)

ABSTRACT

In this chapter, an optimization algorithm that is based on the kinetic energy of gas molecules, namely kinetic gas molecule optimization (KGMO), is introduced. This algorithm has some agents that are gas molecules, which move in the search space; these agents are subject to the kinetic theory of gases, which defines the rules for gas molecule interactions in the model. This algorithm has a good performance in terms of finding the global minima in 23 nonlinear benchmark functions, and the performance is compared with two other benchmark algorithms, namely particle swarm optimization (PSO) and the recently developed high-performance gravitational search algorithm (GSA).

8.1 INTRODUCTION

There are various types of optimization problems, which change with time in the real-world. Typical examples of optimization problems include recognition of moving objects in changing background; data mining; routing in networks, etc. Finding a solution for optimization problems has been studied by many researchers. It is understood that the classical optimization algorithms do not provide suitable solutions in a high-dimensional search space, because of the exponential increase of the problem size. Therefore, traditional techniques such as exhaustive search are not practical in solving these problems (Arkin, 1998; Beni and Wang; 1989; Bonabeau et al., 1999).

DOI: 10.4018/978-1-5225-5580-3.ch008

Many algorithms have been developed for the problem of optimization of objective functions, which is the focus of this chapter. However, they suffer from shortcomings such as getting trapped in local optima, slow convergence and complexity of operations in finding the global optimum point. For example, the standard Particle Swarm Optimization (PSO) algorithm easily gets trapped in the local optima when solving complex multimodal problems (Devi et al., 2011; Idoumghar et al., 2011), while Genetic Algorithm (GA) has no absolute assurance of finding a global optimum. This happens very often when the populations have a large number of subjects. On the other hand, there are various operations such as mutation and crossover in GA that causes the long response time for finding the solution. Evolutionary algorithms (EA) too suffer from slow convergence problem (Devi et al., 2011). In the method proposed in this work, the aim is to gain the best global optima in less number of iterations, thus gaining in convergence time and simplicity of operations.

There are two important aspects in swarm-based heuristic algorithms: exploration and exploitation. The ability of expanding the search space is called exploration, that is an important ability in a swarm-based heuristic algorithm, and the ability of finding the optima around a good solution is called exploitation. Exploring the search space to find new solutions happens in premier iterations in a heuristic search algorithm. This is useful to avoid being trapped in a local optimum. It is necessary that a good balance between exploration and exploitation exists. Extra exploration would cause a pure random search, and extra exploitation would cause trapping in local search (Chun-an, 2008; Rashedi et al. 2009).

Overall, a review of the presented algorithms shows that there is no one superior method for solving optimization problems. The aim of the algorithm proposed in this paper is to optimise the objective functions based on the behaviour of gas molecules. The performance of the proposed Kinetic Gas Molecule Optimization (KGMO) algorithm is evaluated against two optimization algorithms, PSO and the Gravitational Search Algorithm (GSA). GSA is an optimization algorithm based on the law of gravity and mass interactions. It will be shown that the gas molecule behaviour can provide better performance in less number of iterations compared to PSO and GSA benchmark algorithms with high optimization performance.

The next section introduces the fundamentals of the kinetic theory of gas. In Section 8.3, KGMO and its characteristics are described. In Section 8.4, a comparative study is presented, with a demonstration on the experimental

results given in Section 8.5 and a discussion in Section 8.6. This is followed by the concluding remarks in Section 8.7.

8.2 KINETIC THEORY OF GAS MOLECULE

This part provides a brief description of relevant basic concepts of gas molecule laws. Scientists such as Boyle, Charles and Gay-Lussac developed the gas laws based on empirical observations and described the macroscopic behavior of gas molecules, which are the properties that can be directly observed and experienced by a person. The atomic theory of gases states that each substance is composed of a large number of very small particles (molecules or atoms). Basically, all the properties of gases including pressure, volume and temperature, are the consequence of the actions of the molecules making up the gas (Leonard, 2004).

There are five postulates that describe the behavior of molecules in a gas. The kinetic molecular theory of ideal gases is as below (Hopwood and Jeans, 2009):

1. A gas consists of a collection of small particles travelling in straight-line motion. The movement is based on Newton's Laws.
2. The molecules in a gas occupy no volume, which means they are points.
3. Collisions between molecules are perfectly elastic, which means no energy is gained or lost during the collision.
4. There are no attractive or repulsive forces between the molecules.
5. The average kinetic energy of a molecule is $3kT/2$, where T is the absolute temperature and k is the Boltzmann constant having the value 1.38×10^{-23}.

The ideal gas rule is as below (Hopwood and Jeans, 2009):

$$PV = NkT \tag{1}$$

where P is the pressure of gas, V is the volume of the container, and N is the number of particles in the gas.

The energy associated with the movement of an object is called kinetic energy. The faster something moves, the greater its kinetic energy. Therefore, a stationary object has zero kinetic energy. Kinetic energy does not include

other forms of energy, such as thermal or potential energy. Below is the derivation of the kinetic energy equation:

$$\Delta k = Q = F\Delta s = ma\Delta s \tag{2}$$

where Δk is difference of kinetic energy of the gas molecule in the old and new positions, Q is the work energy, F is the Newton force, a is the acceleration, Δs is difference of position of a gas molecule in the unit of time interval and m is the mass of the gas molecule.

Then, from the kinematics equations and rearranging:

$$v^2 = v_0^2 + 2a\Delta s \Rightarrow a\Delta s = \frac{v^2 - v_0^2}{2} \tag{3}$$

where v is the velocity of the gas molecule in the new position, v_0 is the velocity of the gas molecule in the old position and a is the acceleration of the gas molecule.

Combining the two expressions:

$$\Delta k = m\left(\frac{v^2 - v_0^2}{2}\right) \tag{4}$$

and then by expanding Eq. (4), we get:

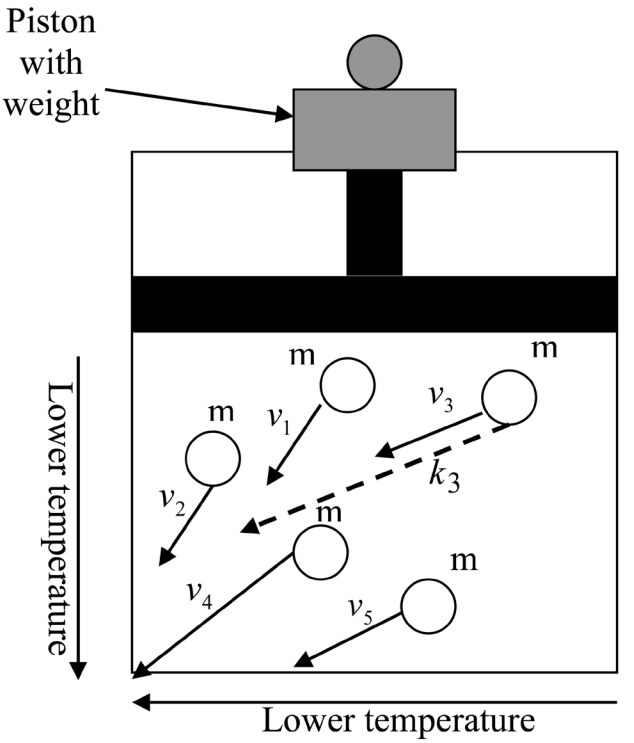
$$\Delta k = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2 \tag{5}$$

Naturally, the kinetic energy of an object at rest should be zero. Thus, an object's kinetic energy is defined mathematically by the following equation:

$$k = \frac{1}{2}mv^2 \tag{6}$$

To illustrate the principle, consider a vertical cylinder closed at the bottom, with a piston at the top, the piston having a weight on it, both piston and weight being supported by the air pressure inside the cylinder. Based on rule of ideal gas (Equation 1), having a constant pressure, and decreasing

Figure 1. In constant pressure, the kinetic energy of gas molecules decrease by decreasing the velocity



the temperature, gas molecules converge in the part of container with lower temperature. The lower the temperature, the less is the motion Equation (6). This is illustrated in Figure 1. If the left lower part of the container has the lowest temperature, then gas molecules converge there after passing time intervals. This principle and Equation (6) are used in the proposed optimization algorithm, as described in the next section.

8.3 KINETIC GAS MOLECULE OPTIMIZATION (KGMO)

In the proposed Kinetic Gas Molecule Optimization (KGMO) algorithm, agents are considered as gas molecules and their performance is measured by their kinetic energy. The gas molecules move in the container until they converge in the part with the lowest temperature and kinetic energy. Gas

molecules attract each other based on weak electrical intermolecular Van Der Waal forces. The electrical force is the result of positive and negative charges in molecules. Comparing to thermal energy this force is weak and has no significant effect on the velocity of molecules.

In KGMO, each gas molecule (agent) has four specifications: position, kinetic energy, velocity and mass of the gas molecule. The kinetic energy of each gas molecule determines its velocity and position. The KGMO could be considered as an isolated system of gas molecules in the container, that explore the whole space to reach the point with the lowest temperature. It is like a small artificial world of gas molecules obeying the kinetic theory of gas molecules.

Now, consider a system with N agents (gas molecules). The position of the i th agent is defined by:

$$X_i = (x_i^1, ..., x_i^d, ...x_i^n), \text{ for } (i = 1, 2, ..., N) \tag{7}$$

where x_i^d represents the position of the i th agent in the d th dimension.

Also, the velocity of the i th agent is presented by:

$$V_i = (v_i^1, ...v_i^d, ..., v_i^n), \text{ for } (i = 1, 2, ..., N) \tag{8}$$

where v_i^d represents the velocity of the i th agent in the d th dimension.

Using Equation (6), the kinetic energy of the i th agent in the d th dimension is defined as:

$$K_i = \frac{1}{2} m V_i^2, \; K_i = (k_i^1, ...k_i^d, ..., k_i^n), \text{ for } (i = 1, 2, ..., N) \tag{9}$$

where k_i^d represents the kinetic energy of the i th agent in the d th dimension.

The velocity equation is defined by equation (10):

$$v_i^d = \frac{x_i^d}{\Delta t} \tag{10}$$

At a unit of time interval (Δt), using the velocity equation, the x_i^d is defined by:

$$x_i^d = v_i^d, \text{ if } \Delta t = 1 \tag{11}$$

From Equation (6) and Equation (11), (Leonard, 2004):

$$k_i^d(t) = \frac{1}{2} m(x_i^d(t))^2 \tag{12}$$

Using Equation (12), the velocity and position are updated by Equation (13) and Equation (14), respectively:

$$x_i^d(t+1) = x_i^d(t) + \sqrt{\frac{2k_i^d(t)}{m}} \tag{13}$$

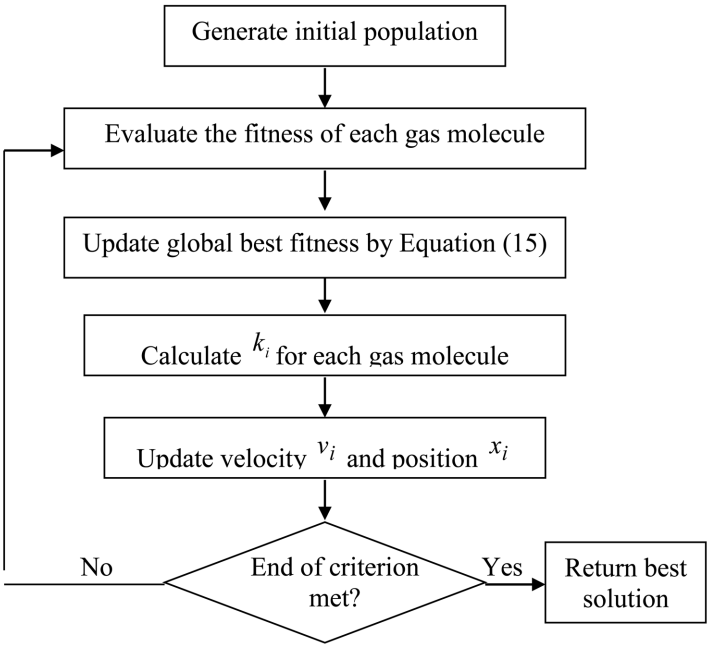
$$\begin{aligned} &v_i^d(t+1) = \\ &Ewv_i^d(t) + c_1 rand_i()(gbest^d - x_i^d(t)) + c_2 rand_i()(Pbest_i^d(t) - x_i^d(t)) \end{aligned} \tag{14}$$

where E is the temperature of gas, which reduces by passing time for converging the molecules, $pbest_i = (pbest_i^1, pbest_i^2, ..., pbest_i^n)$ and $gbest = (gbest^1, gbest^2, ..., gbest^n)$ represent the best previous position of the ith gas molecule and the best pervious position among all the molecules in the container, respectively. w is the inertia weight which reflects the gas molecule resistance to slow its movement. $rand_i$ is a uniform random variable in the interval [0 1], used to provide a randomised characteristic to the search algorithm. c_1 and c_2 are two acceleration constants. m is the mass of each molecule ($0 \prec m \leq 1$). The $\left[-v_{\min}, \quad v_{\max}\right]$ is considered for the gas molecules maximum velocity at advance. If $\left|v_i\right| \succ v_{\max}$, then $\left|v_i\right| = v_{\max}$. The minimum fitness function is found using:

$$pbest_i = f(x_i), \text{ if } f(x_i) \prec f(pbest_i) \tag{15}$$

$$gbest = f(x_i), \text{ if } f(x_i) \prec f(gbest)$$

Figure 2. General steps of KGMO



Each gas molecule tries to modify its position (x_i^d) using the distance between the current position and $pbest_i^d$, and the distance between the current position and $gbest_i$. The principle of KGMO is presented in Figure 2, with the pseudo-code for the algorithm given in Code 1.

```
Code 1. Pseudo-code for KGMO algorithm
For each gas molecule {
          Repeat initialise gas molecules until it satisfies
all the constraints
}

Do {
    For each particle {
        Calculate fitness value
        If the fitness value is better than the best fitness
value (pbest)
        Set current value as the new pbest
      }
     For each particle {

Choose the particle with the best fitness value of all (gbest)
```

```
Calculate the kinetic energy of each gas molecule (Equation
(12)).
Update particle position (Equation (13)) and velocity (Equation
(14))
        }
} while maximum iterations or minimum error criterion not
attained
```

8.4. COMPARATIVE STUDY

In order to examine how the proposed Kinetic Energy of Gas Molecules (KGMO) is situated compared to other heuristic search optimization algorithms, experimental comparison with other algorithms is undertaken. Based on the recent results in (Rashedi et al., 2009), GSA has better performance in finding the optima in 1000 training cycles for most of the benchmark functions in comparison with PSO, Real Genetic Algorithm (RGA) and Central Force Optimization (CFO) algorithms, which are from the newest swarm methods developed by researchers. In addition, PSO has been used as the benchmark algorithm to evaluate the performance of various new optimization algorithms (Chun-an, 2008; Devi et al., 2011; Wolpert and Macready, 1997; Yao et al., 1999; Ying and Pei, 2010; Zhan et al., 2009). Therefore, GSA and PSO are selected among the algorithms to evaluate the performance of KGMO in finding the global optima. The fundamentals of the PSO and GSA algorithms are described below.

8.4.1 PSO Algorithm

As described in chapter 4, PSO updates the population of particles by updating the following 2 equations:

$$x_i^d(t+1) = x_i^d(t) + v_i^d(t+1) \tag{16}$$

$$v_i^d(t+1) = w(t)v_i^d(t) + c_1 rand()(pbest_i^d - x_i^d(t)) + c_2 rand()(gbest^d - x_i^d(t)) \tag{17}$$

$pbest_i = (pbest_i^1, pbest_i^2, ..., pbest_i^n)$ and $gbest = (gbest^1, gbest^2, ..., gbest^n)$ represent the best previous position of the ith particle and among all the

particles in the population, respectively. c_1 and c_2 are two acceleration constants (Venter and Haftka, 2010). The initialisation process for velocity and position of particles is by vector of random numbers in the corresponding range (Yao, 1999). PSO is one of the simplest optimization algorithms in implementation, therefore, in recent years has rapidly progressed and is applied for successful solving many real-world problems (Buczak and Uhrig, 1996; Chun-an, 2008; Devi et al., 2011; Dong et al., 2007; Dorigo, 1996; Eberhart and Shi, 2001). PSO is an iterative population-based algorithm. Hence, some of the evaluation functions measured inefficient computational performance (Farmer et al., 1986). One of the major problems of PSO is trapping in the local optima when solving multimodal problems (Flake, 1999). Therefore, wider application of PSO has restricted because of some of these weaknesses (Chun-an, 2008). Some PSO algorithms are developed to achieve to two major goals of PSO that are accelerating convergence speed and avoiding the local optima (Dorigo et al., 1996; Eberhart and Shi, 2001; Farmer et al., Flake, 1999).

8.4.2 Gravitational Search Algorithm (GSA)

GSA is a population based search algorithm based on the law of gravity and mass interaction (Rashedi et al., 2009). Objects desire to move towards the heavier mass based on the gravitational attraction force between them. There are four specific parameters in GSA: position of the mass in the d th dimension, inertia mass, active gravitational mass and passive gravitational mass.

GSA determines the positions of the mass at specified dimension, with the determined positions reflecting the solution of the problem. It has a randomised initialisation process. The inertia parameter is the resistance of the mass of agents for slowing their movement. The gravitational mass and inertial mass are computed by fitness evolution of the problem. In each iteration, the position of masses is updated. A fixed amount of iterations defines the termination condition of the algorithm. After termination of the algorithm, the position of the mass at specified dimensions of the corresponding agent becomes the global fitness for a particular problem (Rashedi et al., 2009). To describe the GSA, consider a system with *s* masses in which the position of the i th mass is defined as follows:

$$X_i = (x_i^1, ..., x_i^d, ...x_i^n), \text{ for } (i = 1, 2, ..., N) \tag{18}$$

where x_i^d presents the position of the ith agent in the dth dimension.

The mass of each agent is calculated after computing the current population's fitness as follows:

$$M_i(t) = \frac{q_i(t)}{\sum_{j=1}^{n} q_j(t)} \tag{19}$$

$$q_i(t) = \frac{fit_i(t) - worst(t)}{best(t) - worst(t)} \tag{20}$$

where $fit_i(t)$ and $M_i(t)$ represent the fitness value and mass of the agent i at time t, respectively. For a minimisation problem, $worst(t)$ and $best(t)$ are defined as follows:

$$best(t) = \min_{j \in \{1,\ldots,N\}} fit_j(t) \tag{21}$$

$$worst(t) = \max_{j \in \{1,\ldots,N\}} fit_j(t) \tag{22}$$

To compute the acceleration of an agent, total forces from a set of heavier masses applied on an agent should be considered based on the law of gravity (Equation 23), which is followed by calculation of agent acceleration using the law of motion (Equation 24).

$$F_i^d(t) = \sum_{j \in kbest, j \neq i} rand_j G(t) \frac{M_j(t) M_i(t)}{R_{ij(t)} + \varepsilon} (x_j^d(t) - x_i^d(t)) \tag{23}$$

$$a_i^d(t) = \frac{F_i^d(t)}{M_i(t)} = \sum_{j \in kbest, j \neq i} rand_j G(t) \frac{M_j(t)}{R_{ij(t)} + \varepsilon} (x_j^d(t) - x_i^d(t)) \tag{24}$$

The velocity and position of an agent are updated according to Equations (25) and (26), respectively:

$$v_i^d(t+1) = rand_i v_i^d(t) + a_i^d(t) \tag{25}$$

$$x_i^d(t+1) = x_i^d(t) + v_i^d(t+1) \tag{26}$$

where $rand_i$ and $rand_j$ are two uniformly distributed random numbers in the interval [0,1], ε is a small constant, $R_{ij}(t)$ is the Euclidean distance between two agents i and j, defined as $\left\| X_i(t), X_j(t) \right\|_2$, *kbest* is the set of first *K* agents with the best fitness value and biggest mass, which is a function of time, initialised to K_0 at the beginning and decreasing with time. K_0 is set to *N* (total number of agents) and is decreased linearly to 1. Comparing the GSA and the proposed algorithm shows the simplicity of the KGMO in terms of mathematical complexity and reduced number of equations.

8.5 EXPERIMENTAL RESULTS

To evaluate the performance of KGMO, 23 standard benchmark functions from (Yao et al. 1999) are used. Empirical study papers used small number of functions. However, small number of functions causes difficulty in generalising the conclusion. Wolpert and Macready (1997) have shown that no algorithm has the superior performance for all problems under certain assumptions. The benchmark functions used are presented next, followed by the performance comparison of KGMO with PSO and GSA.

8.5.1 Benchmark Functions

The benchmark functions as obtained from (Yao et al., 1999), are shown in Table 1. These sets of functions were designed to test various aspects of algorithms. $f1(x)$ to $f7(x)$ are unimodal functions that have one global minimum with no local minimum. $f8(x)$ to $f13(x)$ are multimodal functions where the number of local minima increases exponentially with the problem dimension. They appear to be the most difficult class of problems for many optimization algorithms. $f14(x)$ to $f23(x)$ are multimodal test functions with

Table 1. Benchmark functions

Test Function	n	s	$f_{\min}$
$f1(x)=\sum_{i=1}^{n}x_i^2$	30	[-100, 100]^n	0
$f2(x)=\sum_{i=1}^{n}\lvert x_i\rvert+\prod_{i=1}^{n}\lvert x_i\rvert$	30	[-10, 10]^n	0
$f3(x)=\sum_{i=1}^{n}(\sum_{j=1}^{i}x_j)^2$	30	[-100, 100]^n	0
$f4(x)=\max_i\{\lvert x_i\rvert,1\le i\le n\}$	30	[-100, 100]^n	0
$f5(x)=\sum_{i=1}^{n-1}[(100(x_{i+1}-x_i^2)^2+(x_i-1)^2]$	30	[-30, 30]^n	0
$f6(x)=\sum_{i=1}^{n}(\lfloor x_i+0.5\rfloor)^2$	30	[-100, 100]^n	0
$f7(x)=\sum_{i=1}^{n}ix_i^4+random[0,1)$	30	[-1.2S, 1.2S]^n	0
$f8(x)=\sum_{i=1}^{n}-x_i\sin(\sqrt{\lvert x_i\rvert})$	30	[-500, 500]^n	-12569.5
$f9(x)=\sum_{i=1}^{n}[x_i^2-10\cos(2\pi x_i)+10]$	30	[-5.12, 5.12]^n	0
$f10(x)=-20\exp\left(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2}\right)-\exp\left(\frac{1}{n}\sum_{i=1}^{n}\cos 2\pi x_i\right)+20+e$	30	[-32, 32]^n	0
$f11(x)=\frac{1}{4000}\sum_{i=1}^{n}x_i^2-\prod_{i=1}^{n}\cos\left(\frac{x_i}{\sqrt{i}}\right)+1$	30	[-600, 600]^n	0

continued on following page

Table 1. Continued

Test Function	n	s	$f_{\min}$
$f12(x) = \frac{\pi}{n}\left\{\begin{array}{l}10\sin(\pi y_1) + \\ \sum_{i=1}^{n-1}(y_i-1)^2[1+10\sin^2(\pi y_{i+1})] + (y_n-1)^2\end{array}\right\} + \sum_{i=1}^{n} u(x_i,10,100,4)$ $y_i = 1 + \frac{x_i+1}{4}$ $u(x_i,a,k,m) = \begin{cases} k(x_i-a)^m, & x_i > a \\ 0, & -a \le x_i \le a \\ k(-x_i-a)^m, & x_i \prec -a \end{cases}$	30	[-50, 50]^n	0
$f13(x) = 0.1\left\{\begin{array}{l}\sin^2(3\pi x_1) + \\ \sum_{i=1}^{n}(x_i-1)^2\left[1+\sin^2(3\pi x_i+1)\right] + \\ (x_n-1)^2[1+\sin^2(2\pi x_n)]\end{array}\right\} + \sum_{i=1}^{n} u(x_i,5,100,4)$	30	[-50, 50]^n	0
$f_{14}(x) = \left(\frac{1}{500} + \sum_{j=1}^{25}\frac{1}{\sum_{i=1}^{2}(x_i - a_{ij})^6}\right)^{-1}$	2	[-65.53, 65.53]	1
$f15(x) = \sum_{i=1}^{11}\left[a_i - \frac{x_1(b_i^2 + b_i x_2)}{b_i^2 + b_i x_3 + x_4}\right]^2$	4	[-5, 5]	0.00030
$f16(x) = 4x_1^2 - 2.1x_1^4 + \frac{1}{3}x_1^6 + x_1x_2 + 4x_2^2 + 4x_2^4$	2	[-5, 5]	-1.03162
$f17(x) = \left(x_2 - \frac{5.1}{4\pi^2}x_1^2 + \frac{5}{\pi}x_1 - 6\right)^2 + 10\left(1 - \frac{1}{8\pi}\right)\cos x_1 + 10$	2	[-5, 10] × [0, 15]	0.398

continued on following page

Table 1. Continued

Test Function	n	s	$f_{\min}$
$f18(x) = \begin{bmatrix} 1+(x_1+x_2+1)^2(19-14x_1+3x_1^2-14x_2+ \\ 6x_1x_2+3x_2^2) \end{bmatrix} \times \begin{bmatrix} 30+\left(2x_1-3x_2\right)^2 \times \\ (18-32x_1+12x_1^2+48x_2-36x_1x_2+27x_2^2) \end{bmatrix}$	2	[-5, 5]	3
$f19(x) = -\sum_{i=1}^{4} c_i \exp\left(-\sum_{j=1}^{3} a_{ij}(x_j - p_{ij})^2\right)$	3	[0, 1]	-3.86
$f20(x) = -\sum_{i=1}^{4} c_i \exp\left(-\sum_{j=1}^{6} a_{ij}(x_j - p_{ij})^2\right)$	6	[0, 1]	-3.32
$f21(x) = -\sum_{i=1}^{5} \left[(x-a_i)(x-a_i)^T + c_i\right]^{-1}$	4	[0, 10]	-10
$f22(x) = -\sum_{i=1}^{7} \left[(x-a_i)(x-a_i)^T + c_i\right]^{-1}$	4	[0, 10]	-10
$f23(x) = -\sum_{i=1}^{10} \left[(x-a_i)(x-a_i)^T + c_i\right]^{-1}$	4	[0, 10]	-10

fixed dimension which have few local minima. The detailed description of the parameters is provided in Table 2. The best algorithm is one that finds the nearest optima to the actual global minimum.

8.5.2 Results Comparison

The results of evaluation of KGMO against GSA and PSO for the three types of functions: unimodal, multimodal and multimodal test functions with fix dimension, is described in this section. The minimum of each function is the average of the best global minimum obtained by each algorithm over 50 runs. In all cases, the population size is set to 50 *(N = 50)*. For PSO, $c_1 = c_2 = 2$ and inertia factor (w) is decreased linearly from 0.85 to 0.2, as per the default values.

Table 2. The benchmark functions parameter (a) a_{ij} in $f14$.

$$(a_{ij}) = \begin{pmatrix} 32, -16, 0, 16, 32, -32, -16, 0, 16, 32, -32, ..., 0, 16, 32, -32, -16, 0, 16, 32, \\ -32, -32, -32, -32, -32, -16, ..., 0, 16,16, 16, 16, 16, 32, 32, 32, 32, 32, \end{pmatrix}$$

(b) a_i and b_i in $f15$.

i	1	2	3	4	5	6	7	8	9	10	11
a_i	0.1957	0.1947	0.1735	0.16	0.0844	0.0627	0.0456	0.0342	0.0323	0.0235	0.0246
b_i^{-1}	0.25	0.5	1	2	4	6	8	10	12	14	16

(c) a_{ij} and c_i in $f19$.

i	$a_{ij}, j = 1,2,3$			c_i
1	3	10	30	1
2	0.1	10	35	1.2
3	3	10	30	3
4	0.1	10	30	3.2

(d) p_{ij} in $f19$.

i	$p_{ij}, j = 1,2,3$		
1	0.3689	0.1170	0.2673
2	0.4699	0.4387	0.7470
3	0.1091	0.8732	0.5547
4	0.03815	0.5743	0.8828

(e) a_{ij} and c_i in $f20$.

i	$a_{ij}, j = 1,2,3,4,5,6$						c_i
1	10	3	17	3.5	1.7	8	1
2	0.05	10	17	0.1	8	14	1.2
3	3	3.5	1.7	10	17	8	3
4	17	8	0.05	10	0.1	14	3.2

continued on following page

Table 1. Continued

(f) p_{ij} in $f20$.						
i	$p_{ij}, j = 1,2,3,4,5,6$					
1	0.131	0.169	0.556	0.012	0.828	0.588
2	0.232	0.413	0.830	0.373	0.100	0.999
3	0.234	0.141	0.352	0.288	0.304	0.665
4	0.404	0.882	0.873	0.574	0.109	0.038

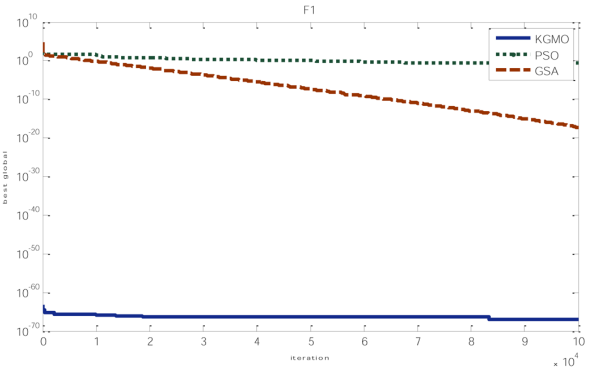
(g) a_{ij} and c_i in $f21$, $f22$ and $f23$.					
i	$a_{ij}, j = 1,2,3,4$				c_i
1	4	4	4	4	0.1
2	1	1	1	1	0.2
3	8	8	8	8	0.2
4	6	6	6	6	0.4
5	3	7	3	7	0.4
6	2	9	2	9	0.6
7	5	5	3	3	0.3
8	8	1	8	1	0.7
9	6	2	6	2	0.5
10	7	3.6	7	3.6	0.5

Based on the initialisation in (Rashedi et al. 2009) for GSA, G_0, the initial gravitational constant, is set to 100 and α is set to 20, K_0 is set to *N* which is 50 and T is the total number of iterations. In KGMO, $c_1 = 1$, $c_2 = 3$, $m = 1$ and inertia factor (w) is decreased linearly from 0.85 to 0.2. The *E* decreased linearly from 0.95 to 0.1. In (Rashedi et al., 2009), it was found that GSA already provides satisfying results for finding the global minimum of 23 functions at 1000 iterations. This work attempts to formulate an improved algorithm that is able to find the optima in the least number of training cycles.

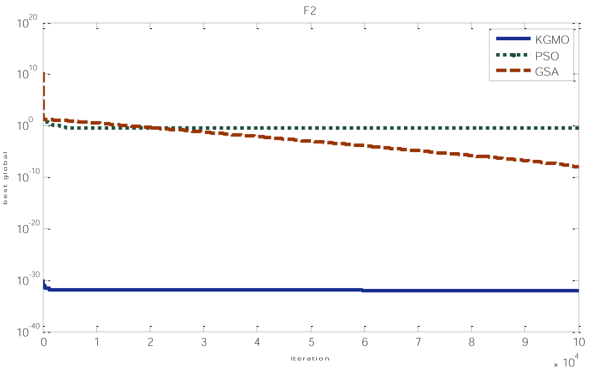
8.5.2.1 Unimodal High-dimensional Functions

Functions $f1$ to $f7$ are unimodal functions for which the convergence rate is more important as current optimization algorithms present good final

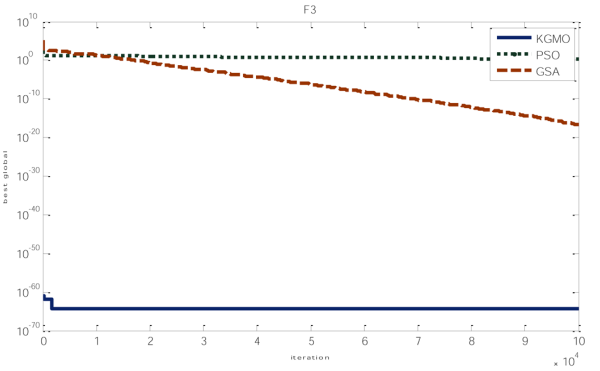
Figure 3. Average performance for minimisation of unimodal functions over 50 runs of 100000 iterations



(a) $f1(x)$



(b) $f2(x)$



(c) $f3(x)$

results. Therefore, the aim is to obtain the best global minimum in the least number of required iterations [19].

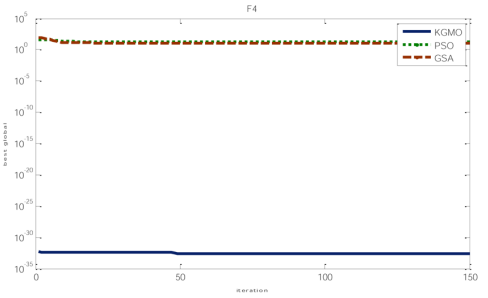
The nearest *Average Gbest* to $f(opt)$ determines which algorithm works better. Figure 3 presents the convergence of KGMO, GSA and PSO for functions $f1(x)$, $f2(x)$ and $f3(x)$ in 100000 iterations. The figures show that KGMO tends to find the global minimum in less than 150 iterations and the optima in the last iteration is very close to the actual minimum value. It is observed that that PSO and GSA could not obtain the KGMO result even after 100000 iterations. In view of time efficiency, 150 is selected as the number of iterations for all tests and experiments in the remainder of the evaluations.

Figure 4 shows the convergence of $f4(x)$ to $f7(x)$ over 50 runs of 150 iterations. It is obvious that KGMO tends to find the best global minimum in the less than 150 iterations with better results compared to PSO and GSA. Table 3 presents the results of the comparison of KGMO with PSO and GSA for finding the minimum value of $f1(x)$ to $f7(x)$ in 150 iterations. The *Average Gbest* and *STD* are the average global best value and standard deviation, respectively, over 50 runs that are used for comparing the performance of the algorithms. In each of the 50 runs, the fitness function values are stored in an array. After finishing the last run, the average and the median of the generated array report the *Average cost* and the *Median cost,* respectively, to show how all molecules and particles converged toward the optimum. $f_{\min}$ is the minimum value of each function. The *Average Gbest* in Table 3 for PSO, GSA and KGMO shows that the nearest result to the actual global minimum. In addition, comparing the STD of all three algorithms presents that the KGMO has the smallest STD for all functions $f1(x)$ to $f7(x)$, evidence of its robustness and efficiency in finding the global optima for unimodal functions. The last row of the table shows the mean squared error (MSE) that is calculated by:

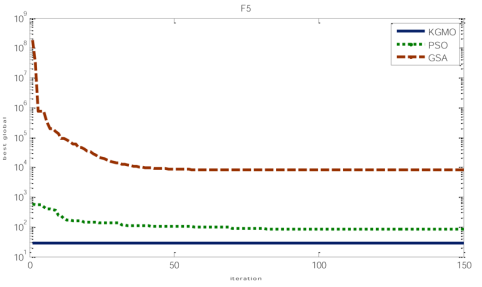
$$MSE = \frac{\sum_{i=1}^{n}\left(Gbest_i - f(opt)_i\right)^2}{n} \tag{27}$$

Gbest is the Average Gbest and *f(opt)* is the target in each row. For unimodal high-dimensional functions, $i = (1,2,..,7)$ and $n = 7$; for multimodal high-dimensional functions $i = (8,9,...13)$ and $n = 6$; and for multimodal high-dimensional functions with fixed dimensions $i = (14,15,...,23)$ and $n = 10$.

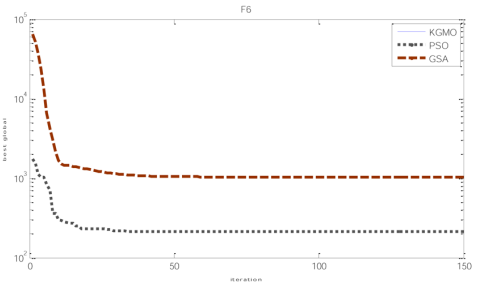
Figure 4. Average performance for minimisation of unimodal functions over 50 runs of 100000 iterations



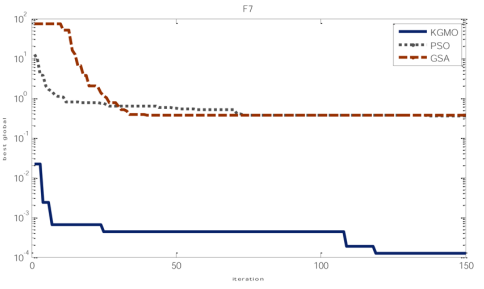
(a) $f4(x)$



(b) $f5(x)$



(c) $f6(x)$



(d) $f7(x)$

Table 3. Average minimisation result of unimodal benchmark functions in Table 1 over 50 runs of 150 iterations

	PSO	GSA	KGMO	$f(opt)$
$f1(x)$	**Average Gbest:** 87.0678 **Average cost:** 2.2124 $\times 10^{3}$ **Median cost:** 87.2559 **STD:** 21.5291	**Average Gbest:** 1111 **Average cost:** 2.3437 $\times 10^{3}$ **Median cost:** 1.1115 $\times 10^{3}$ **STD:** 7.7814 $\times 10^{3}$	**Average Gbest:** 1.5177 $\times 10^{-67}$ **Average cost:** 1.2171 $\times 10^{-63}$ **Median cost:** 3.8486 $\times 10^{-66}$ **STD:** 1.3704 $\times 10^{-67}$	0
$f2(x)$	**Average Gbest:** 13.6616 **Average cost:** 13.9796 **Median cost:** 13.6642 **STD:** 4.7100	**Average Gbest:** 1 $\times 10^{7}$ **Average cost:** 1.8637 $\times 10^{9}$ **Median cost:** 0.4029 **STD:** 2.2826 $\times 10^{10}$	**Average Gbest:** 6.6717 $\times 10^{-32}$ **Average cost:** 4.3344 $\times 10^{-30}$ **Median cost:** 3.3797 $\times 10^{-30}$ **STD:** 8.9356 $\times 10^{-32}$	0
$f3(x)$	**Average Gbest:** 15.5011 **Average cost:** 9451.1 **Median cost:** 16.0331 **STD:** 27.3169	**Average Gbest:** 48000 **Average cost:** 7888.1 **Median cost:** 4803.4 **STD:** 8.3495 $\times 10^{7}$	**Average Gbest:** 2.2503 $\times 10^{-61}$ **Average cost:** 2.4722 $\times 10^{-61}$ **Median cost:** 1.5414 $\times 10^{-58}$ **STD:** 4.2420 $\times 10^{-62}$	0
$f4(x)$	**Average Gbest:** 19.7695 **Average cost:** 19.7700 **Median cost:** 19.7695 **STD:** 6.3284	**Average Gbest:** 11.6925 **Average cost:** 11.6925 **Median cost:** 11.6925 **STD:** 4.8745	**Average Gbest:** 1.4984 $\times 10^{-33}$ **Average cost:** 1.4485 $\times 10^{-31}$ **Median cost:** 1.0216 $\times 10^{-31}$ **STD:** 1.1348 $\times 10^{-33}$	0
$f5(x)$	**Average Gbest:** 82.8135 **Average cost:** 73.2547 **Median cost:** 28.8522 **STD:** 14.2206 $\times 10^{1}$	**Average Gbest:** 10 $\times 10^{4}$ **Average cost:** 2.6043 $\times 10^{5}$ **Median cost:** 26.0451 **STD:** 7.1841 $\times 10^{6}$	**Average Gbest:** 24.8522 **Average cost:** 24.8522 **Median cost:** 24.8522 **STD:** 2.2277 $\times 10^{-4}$	0

continued on following page

Table 3. Continued

	PSO	GSA	KGMO	$f(opt)$
$f6(x)$	**Average Gbest:** 336 **Average cost:** 38.022 $\times 10^{1}$ **Median cost:** 336 **STD:** 11.5497	**Average Gbest:** 845 **Average cost:** 23.104 $\times 10^{2}$ **Median cost:** 79.3436 **STD:** 2.5766 $\times 10^{3}$	**Average Gbest:** 0 **Average cost:** 0 **Median cost:** 0 **STD:** 0	0
$f7(x)$	**Average Gbest:** 0.3607 **Average cost:** 18.6847 **Median cost:** 0.8032 **STD:** 10.8253	**Average Gbest:** 0.35185 **Average cost:** 0.7 **Median cost:** 0.7 **STD:** 23.6470	**Average Gbest:** 0.1669 $\times 10^{-4}$ **Average cost:** 0.1621 $\times 10^{-3}$ **Median cost:** 0.1022 $\times 10^{-3}$ **STD:** 1.1119 $\times 10^{-3}$	0
MSE	1.7319 $\times 10^{4}$	1.4286 $\times 10^{17}$	1.78 $\times 10^{-3}$	

The MSE in the last row of Table 3 shows that in 150 iterations, KGMO has decreased the MSE by 10^7 and 10^{20} times compared to PSO and GSA, respectively. Based on Rashedi et. al (2009), for unimodal and multimodal functions with high dimensions, GSA can provide good results in 1000 iterations. It has also good performance in 500 iterations for multimodal functions with fixed dimensions.

8.5.2.2 Multimodal High-Dimensional Functions

Functions in this group are complex as there is the risk of being trapped in many local optima when finding the global optima. Therefore, a good optimization function is one that can find the optima close to the actual global minimum. On the other hand, one disadvantage of previous algorithms in solving some of the multimodal optimization problems is their slow convergence to a good near-optimum (Yao et al., 1999). The aim of this work for the multimodal high-dimensional functions is to escape to trap in local minimum with higher speed than current algorithms. Table 4 presents the comparison of KGMO with GSA and PSO for the multi high-dimensional functions.

Table 4. Average minimisation result of multimodal benchmark functions in Table 1 over 50 runs of 150 iterations

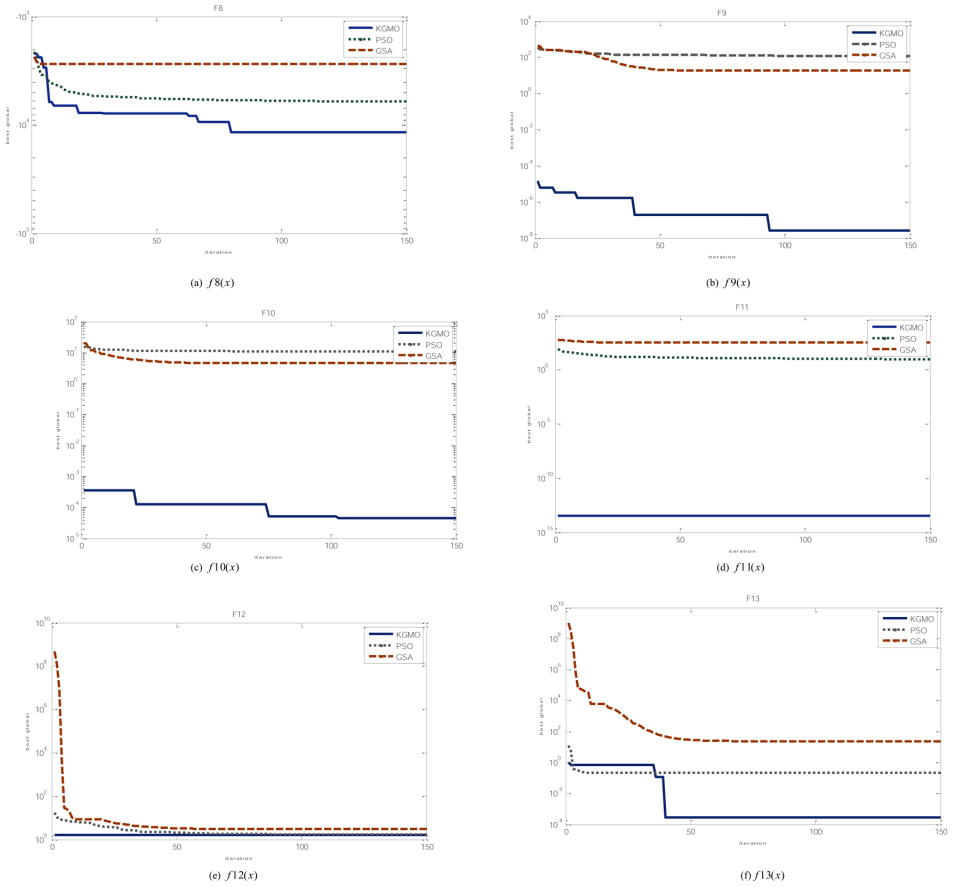
	PSO	**GSA**	**KGMO**	$f(opt)$
$f8(x)$	**Average Gbest:** -6751.8 **Average cost:** -6470 **Median cost:** -6649.3 **STD:** 76.6921 $\times 10^{1}$	**Average Gbest:** -2268.5 **Average cost:** 33.9036 **Median cost:** 5.3449 **STD:** 16.3120 $\times 10^{1}$	**Average Gbest:** -11902 **Average cost:** -24.2020 $\times 10^{1}$ **Median cost:** 46.5653 $\times 10^{1}$ **STD:** 39.7420 $\times 10^{1}$	-12569
$f9(x)$	**Average Gbest:** 11.2384 $\times 10^{1}$ **Average cost:** 14.1680 $\times 10^{1}$ **Median cost:** 11.2622 $\times 10^{1}$ **STD:** 20.2530	**Average Gbest:** 31.5253 **Average cost:** 31.5259 **Median cost:** 31.5259 **STD:** 76.9931	**Average Gbest:** 5.7085 $\times 10^{-8}$ **Average cost:** 4.0373 $\times 10^{-4}$ **Median cost:** 3.0473 $\times 10^{-4}$ **STD:** 5.9230 $\times 10^{-8}$	0
$f10(x)$	**Average Gbest:** 9.2266 **Average cost:** 9.3730 **Median cost:** 9.2375 **STD:** 4.1164	**Average Gbest:** 5.4014 **Average cost:** 5.4014 **Median cost:** 5.4014 **STD:** 2.6500	**Average Gbest:** 1.5637 $\times 10^{-4}$ **Average cost:** 0.0055 **Median cost:** 0.0045 **STD:** 4.5806 $\times 10^{-5}$	0
$f11(x)$	**Average Gbest: 6.3108 Average cost: 7.5496 Median cost: 6.3485 STD: 1.6864**	**Average Gbest: 32.3051** $\times 10^{1}$ **Average cost: 49.6143** $\times 10^{1}$ **Median cost: 50.5975** $\times 10^{1}$ **STD: 44.7624**	**Average Gbest: 1.1102** $\times 10^{-14}$ **Average cost: 5.1434** $\times 10^{-10}$ **Median cost: 1.2084** $\times 10^{-10}$ **STD: 2.0322** $\times 10^{-8}$	0
$f12(x)$	**Average Gbest: 2.9219 Average cost: 9.7624** $\times 10^{6}$ **Median cost: 5.9337 STD: 4.6058**	**Average Gbest: 10** $\times 10^{3}$ **Average cost: 10** $\times 10^{4}$ **Median cost: 10** $\times 10^{4}$ **STD: 4.7665** $\times 10^{7}$	**Average Gbest: 1.6699 Average cost: 1.6741 Median cost: 1.6767 STD: 6.4187** $\times 10^{-5}$	0

continued on following page

Table 4. Continued

	PSO	GSA	KGMO	$f(opt)$
$f13(x)$	**Average Gbest:** **0.0974** **Average cost:** **0.5679** **Median cost:** **0.0974** **STD: 0.8491**	**Average Gbest:** **10** $\times 10^{1}$ **Average cost:** **10** $\times 10^{1}$ **Median cost:** **10** $\times 10^{1}$ **STD: 2.0735** $\times 10^{1}$	**Average Gbest:** **3.4555** $\times 10^{-4}$ **Average cost:** **3.2414** $\times 10^{-2}$ **Median cost:** **1.6023** $\times 10^{-2}$ **STD: 3.5447** $\times 10^{-4}$	0
MSE	**5.6430** $\times 10^{6}$	**2.0110** $\times 10^{4}$	**11.1445** $\times 10^{1}$	

Figure 5. Average performance for minimisation of high-dimensional multimodal functions over 50 runs of 150 iterations



As evaluated in Table 4, KGMO performed better in determining the global minimum of all the multimodal functions. Results show that KGMO could obtain very good performance for *Average Gbest* in all multimodal functions with high-dimensions. It has smaller STD compared with PSO and GSA. The last row of Table 4 shows that KGMO has decreased the MSE by 10^5 and 10^3 times compared to PSO and GSA, respectively. As before, GSA could not obtain good results for $f8(x)$. For further illustration, Figure 5 presents the convergence of KGMO, GSA and PSO for $f8(x)$ to $f13(x)$ in 150 iterations. The figures support that KGMO performs better than PSO and GSA. A further attractive property of KGMO is its fast convergence, as confirmed by the figures.

Table 5. Average minimisation result of multimodal fixed dimension benchmark functions in Table 1 over 50 runs of 150 iterations

	PSO	**GSA**	**KGMO**	$f(opt)$
$f14(x)$	**Average Gbest:** 1.5712 **Average cost:** 0.9926 **Median cost:** 0.7301 **STD:** 0.3826	**Average Gbest:** 3.9731 **Average cost:** 13.6186 **Median cost:** 13.6186 **STD:** 0.8365	**Average Gbest:** 1.5818 **Average cost:** 7.0274 **Median cost:** 2.4728 **STD:** 3.8411×10^{-15}	1
$f15(x)$	**Average Gbest:** 0.0016 **Average cost:** 0.0016 **Median cost:** 0.0016 **STD:** 0.0098	**Average Gbest:** 0.0069 **Average cost:** 100 **Median cost:** 0.0220 **STD:** 0.0087	**Average Gbest:** 0.0012 **Average cost:** 12.9248 **Median cost:** 8.0856 **STD:** 0.0278	0.0003075
$f16(x)$	**Average Gbest:** -1.0316 **Average cost:** -1.0316 **Median cost:** -1.0316 **STD:** 1.5701×10^{-16}	**Average Gbest:** -1.0316 **Average cost:** -1 **Median cost:** -1 **STD:** 0.2621	**Average Gbest:** -1.0113 **Average cost:** -0.9835 **Median cost:** -0.9835 **STD:** 0.4468	-1.03
$f17(x)$	**Average Gbest:** 0.3979 **Average cost:** 0.3979 **Median cost:** 0.3979 **STD:** 0.7297	**Average Gbest:** 0.3979 **Average cost:** 0.3979 **Median cost:** 0.3979 **STD:** 0.0638	**Average Gbest:** 0.4072 **Average cost:** 34.1087 **Median cost:** 37.1256 **STD:** 0.2527	0.398

continued on following page

Table 5. Continued

	PSO	**GSA**	**KGMO**	$f(opt)$
$f18(x)$	**Average Gbest:** 3.0000 **Average cost:** 3.0000 **Median cost:** 3.0000 **STD:** 4.0782×10^{-15}	**Average Gbest:** 3.0000 **Average cost:** 1 **Median cost:** 1 **STD:** 1.3457	**Average Gbest:** 6.0644 **Average cost:** 4.4388 **Median cost:** 5.0212 **STD:** 1.9212	3
$f19(x)$	**Average Gbest:** -3.8628 **Average cost:** -3.8628 **Median cost:** -3.8628 **STD:** 2.5640×10^{-16}	**Average Gbest:** -3.8628 **Average cost:** -0.3351 **Median cost:** -0.3351 **STD:** 0.1459	**Average Gbest:** -3.6267 **Average cost:** -0.4679 **Median cost:** -0.3005 **STD:** 0.0371	-3.86
$f20(x)$	**Average Gbest:** -3.2031 **Average cost:** -3.2031 **Median cost:** -3.2031 **STD:** 0.0627	**Average Gbest:** -3.2031 **Average cost:** -0.7865 **Median cost:** -0.7865 **STD:** 0.0802	**Average Gbest:** -2.9012 **Average cost:** -0.0193 **Median cost:** -3.0415×10^{-4} **STD:** 0.3584	-3.32
$f21(x)$	**Average Gbest:** -2.6552 **Average cost:** -2.6552 **Median cost:** -2.0552 **STD:** 3.7474	**Average Gbest:** -5.0552 **Average cost:** -5.0552 **Median cost:** -5.0552 **STD:** 1.3595	**Average Gbest:** -10.1532 **Average cost:** -2.3266 **Median cost:** -2.1440 **STD:** 2.0656×10^{-4}	-10
$f22(x)$	**Average Gbest:** -10.4023 **Average cost:** -10.4029 **Median cost:** -10.4029 **STD:** 2.6371	**Average Gbest:** -5.0877 **Average cost:** -5.0877 **Median cost:** -5.0877 **STD:** 1.3069	**Average Gbest:** -10.4023 **Average cost:** -1.7011 **Median cost:** -2.1815 **STD:** 3.2995×10^{-4}	-10
$f23(x)$	**Average Gbest:** -10.5364 **Average cost:** -10.5364 **Median cost:** -10.5364 **STD:** 4.5749	**Average Gbest:** -5.1285 **Average cost:** -5.1285 **Median cost:** -5.1285 **STD:** 3.1853	**Average Gbest:** -10.5352 **Average cost:** -1.9536 **Median cost:** -2.2269 **STD:** 2.5275×10^{-4}	-10
MSE	5.4482	5.9169	1.0431	

8.5.2.3 Multimodal Test Functions With Fixed Dimension

The third group of test functions is the set of multimodal functions with fixed dimension. Functions in this group have less number of local optima as compared to those in the previous section. Table 5 shows the performance of the algorithms in finding the minima of $f14(x)$ to $f23(x)$.

From Table 5, comparing PSO and KGMO shows that except for $f18(x)$, $f19(x)$ and $f20(x)$ where PSO performed better, KGMO had good performance in all functions and there was no significant difference in the performance of PSO and KGMO. The very small difference in the performance of PSO and KGMO for $f14(x), f16(x), f17(x)$, $f22(x)$ and $f23(x)$ can be ignored because of the randomness specification of the algorithms. For $f15(x)$ and $f21(x)$ KGMO performed better than PSO. Comparing GSA and KGMO shows that for $f16(x)$, $f17(x)$, $f18(x)$, $f19(x)$ and $f20(x)$, GSA performed better. For other functions, KGMO performed better than GSA.

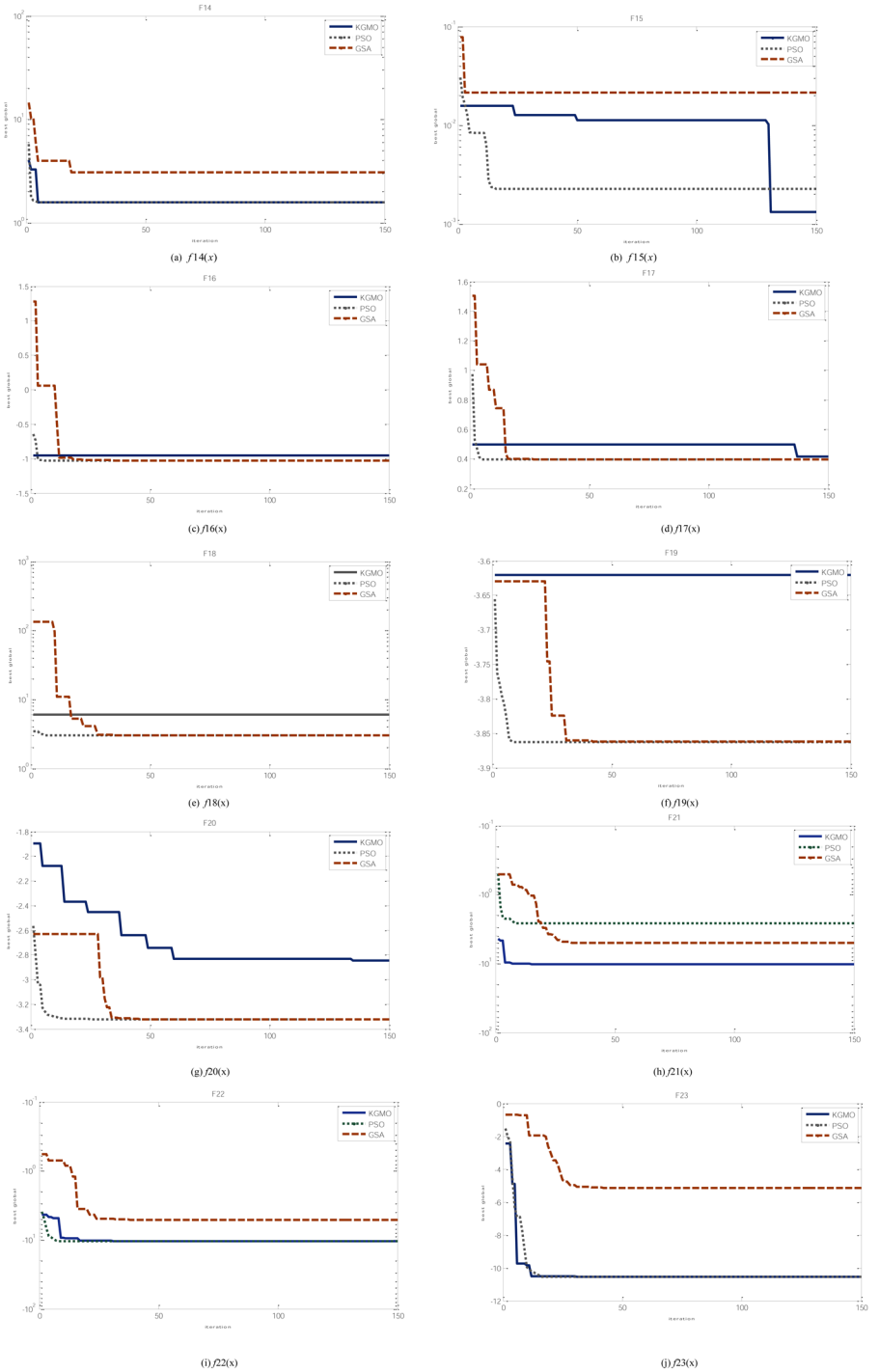
The obtained STD with PSO for functions $f15(x), f16(x), f18(x)$ and $f20(x)$ is smaller than KGMO. However, as the values are not significantly different, PSO and KGMO are of approximately the same in robustness for finding the global optima of the benchmark functions in this group.

The MSE in the last row proves that KGMO has the better overall performance with smaller MSE of around 0.1 compared to about 5 with PSO and GSA. Figure 6 presents the comparison of PSO and GSA with KGMO for functions $f14(x)$ and $f23(x)$. The results support the good performance of KGMO for obtaining the optimal minimum in less than 150 iterations.

8.6 VALIDATION

The Wilcoxon rank sum test is a non-parametric test for assessing the significance difference between two samples of independent observations (Wilcoxon, 1945). The Wilcoxon rank sum test for equal medians is provided for comparing the differentiation of each algorithm *Average Gbest* with $f(opt)$. Table 6 shows the results of Wilcoxon rank sum test with $\alpha = 0.05$ [24]. The results obtained for PSO (*P=0.0068*) and GSA (*P=0.0052*) are smaller than 0.05, showing that the *Average Gbest* for PSO and GSA are significantly different from $f(opt)$. However, the obtained *P=0.0505* for KGMO, which is greater than 0.05, shows that the calculated *Average Gbest* is not significantly different from $f(opt)$.

Figure 6. Average performance for minimisation of multimodal modal functions with fixed dimensions over 50 runs of 150 iterations



(a) $f14(x)$ (b) $f15(x)$

(c) $f16(x)$ (d) $f17(x)$

(e) $f18(x)$ (f) $f19(x)$

(g) $f20(x)$ (h) $f21(x)$

(i) $f22(x)$ (j) $f23(x)$

Table 6. Results of average ranking test and Wilcoxon rank sum test for equal medians for Average Gbest and $f(opt)$

	PSO	**GSA**	**KGMO**	$f(opt)$
$f1(x)$	87.0678	1111	1.52×10^{-64}	0
Rank	2	1	3	
$f2(x)$	13.6616	10000000	6.67×10^{-32}	0
Rank	2	1	3	
$f3(x)$	15.5011	48000	2.25×10^{-61}	0
Rank	2	1	3	
$f4(x)$	19.7695	10.6925	1.50×10^{-33}	0
Rank	2	1	3	
$f5(x)$	94.6729	10000	28.8522	0
Rank	2	1	3	
$f6(x)$	336	845	0	0
Rank	2	1	3	
$f7(x)$	0.0489	0.2437	1.67×10^{-8}	0
Rank	2	1	3	
$f8(x)$	-6751.8	-2268.5	-11902	-12569
Rank	2	1	3	
$f9(x)$	112.3839	31.5253	5.71×10^{-8}	0
Rank	1	2	3	
$f10(x)$	8.2266	5.4014	1.56×10^{-4}	0
Rank	1	2	3	
$f11(x)$	6.3108	323.0513	1.11×10^{-14}	0
Rank	2	1	3	
$f12(x)$	5.9219	10000	1.669	0
Rank	2	1	3	

continued on following page

Table 6. Continued

	PSO	**GSA**	**KGMO**	$f(opt)$
$f13(x)$	0.0974	100000	3.46×10^{-7}	0
Rank	2	1	3	
$f14(x)$	1.2012	3.9731	1.3718	1
Rank	3	1	2	
$f15(x)$	0.0016	0.0069	0.0012	0.0003075
Rank	1	2	3	
$f16(x)$	-1.0316	-1.0316	-1.0113	-1.03
Rank	3	3	1	
$f17(x)$	0.3979	0.3979	0.4072	0.398
Rank	3	3	2	
$f18(x)$	3	3	6.0644	3
Rank	3	3	2	
$f19(x)$	-3.8628	-3.362	-3.8467	-3.86
Rank	3	1	2	
$f20(x)$	-3.2031	-2.2192	-2.9012	-3.32
Rank	3	1	2	
$f21(x)$	-2.6552	-5.0552	-10.1532	-10
Rank	1	2	3	
$f22(x)$	-10.4023	-5.0877	-10.4023	-10
Rank	3	1	3	
$f23(x)$	-10.5364	-10.5364	-10.5352	-10
Rank	3	3	3	
P value	0.0068	0.0052	**0.0505**	
Average Rank	1.9133	1.5652	**2.6956**	

Another test for measuring the performance of KGMO is the Average Ranking (AR) method, in which one rank is assigned to each algorithm for all functions based on the obtained results and finally the ranks are averaged (Wilcoxon, 1945). In Table 6, the best rank (i.e. 3) is assigned to the algorithm with the best *Average Gbest* that is nearest to the actual minimum, rank 2 is assigned to the second algorithm in finding the optima and finally rank 1 to the worst algorithm. If algorithms have the same best and worst *Average Gbest*, rank 3 and 1, respectively, are assigned to them. As Table 6 presents, the average ranking for the obtained results for KGMO is 2.69, which is higher than 1.91 and 1.56 for PSO and GSA, respectively.

Table 7. The average of MSE for 3 models of benchmark functions

	PSO	**GSA**	**KGMO**
Unimodal functions	47.22682	83.8330×10^4	2.40
Multimodal high dimensional functions	59.5014×10^1	12.066×10^3	66.9
Multimodal functions with fixed dimensions	5.445	5.9168	0.1337
Average MSE	21.5895×10^1	28.3467×10^4	**23.14**

Figure 7. MSE of tested algorithms for each type of benchmark functions (Tables 3, 4, 5)

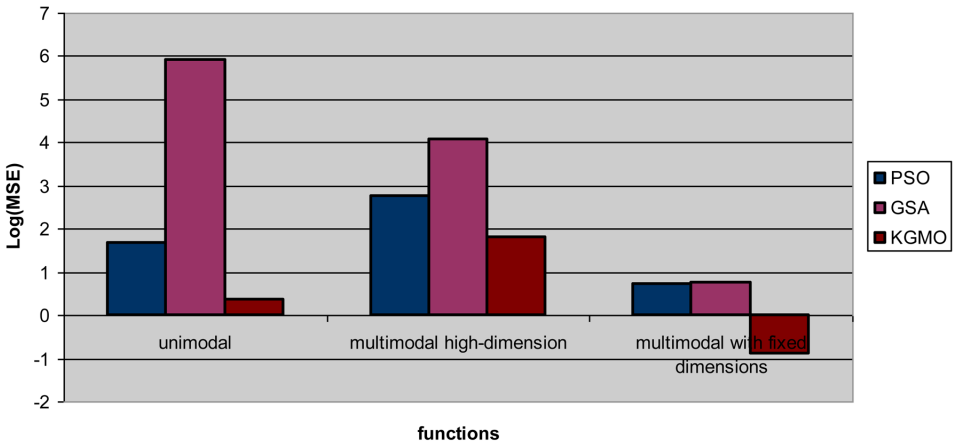


Figure 8. Average MSE of tested algorithms

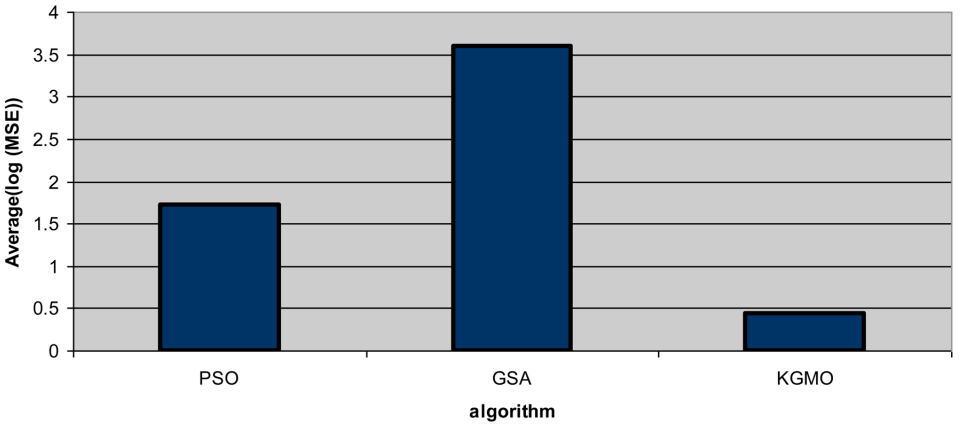


Table 8. Minimisation result of benchmark functions for PSO, GSA and KGMO. Maximum number of iterations for unimodal functions and multimodal high dimensional functions is 1000 and for multimodal functions with fixed dimensions is 500. KGMO results are in 150 training cycles for all functions.

	PSO	**GSA**	**KGMO**	$f(opt)$
$f1(x)$	1.22×10^{-8}	2.12×10^{-17}	1.5177×10^{-67}	0
$f2(x)$	10.3311	2.27×10^{-8}	6.6717×10^{-32}	0
$f3(x)$	12.2221	238.99	2.2503×10^{-61}	0
$f4(x)$	6.9021	3.20×10^{-9}	1.4984×10^{-33}	0
$f5(x)$	52.6255	26.13	24.8522	0
$f6(x)$	0	0	0	0
$f7(x)$	0.0434	0.0487	0.1669×10^{-7}	0
$f8(x)$	-9660.9	-2482.8	-11902	-12569.5
$f9(x)$	93.3482	17.9093	5.7085×10^{-8}	0

continued on following page

Table 8. Continued

	PSO	GSA	KGMO	$f(opt)$
$f10(x)$	6.1121	3.49×10^{-3}	1.5637×10^{-4}	0
$f11(x)$	2.6732	2.4271	1.1102×10^{-14}	0
$f12(x)$	4.7490	1×10^{-18}	1.6690	0
$f13(x)$	5.52×10^{-3}	1×10^{-33}	3.4555×10^{-7}	0
$f14(x)$	0.7301	1.9923	1.3718	1
$f15(x)$	5.33×10^{-3}	0.0084	0.0012	0.00030
$f16(x)$	-1.0316	-1.0316	-1.0113	-1.03162
$f17(x)$	0.3980	0.3979	0.4072	0.398
$f18(x)$	3.00	3.00	3.7423	3
$f19(x)$	-3.86	-3.8347	-3.8467	-3.86
$f20(x)$	-3.3220	-1.3039	-2.9012	-3.32
$f21(x)$	-10.1532	-5.0552	-10.1532	-10
$f22(x)$	-10.4023	-5.0887	-10.4023	-10
$f23(x)$	-10.5364	-10.5364	-10.5352	-10

In addition, the last rows of Tables 3, 4, and 5 for the three types of benchmark functions are compiled in Table 8 and the *Average MSE* for each column is calculated. For clarity and analysis, the graphical representation of the results in Table 7 is also shown in logarithmic representation in Figure 7 and Figure 8. As Figure 7 presents, the KGMO achieves significantly less errors for the three types of functions compared to PSO and GSA. The results in Table 7 and Figure 8 shows that the *Average MSE* for KGMO is around 23, a decrease of 10^1 and 10^4 times compared to PSO and GSA, respectively.

8.6.1 More Results

Table 8 provides the results of global minimum obtained by PSO and GSA in 1000 iterations over 50 runs. It proves that KGMO provides better global minimum than PSO and GSA even in 1000 iterations for the first and second group of functions (i.e. unimodal and multimodal high dimensional functions), which indicates that for unimodal functions, the speed of KGMO is at least 600% faster than PSO and GSA in terms of running iterations to convergence. It also shows that KGMO provides better global minimum than PSO and GSA even in 1000 iterations for all multimodal high dimensional functions, which means that KGMO provides better results at least 600% faster than PSO and GSA in terms of running iterations.

Comparing the results in Table 8 with the results of KGMO in for multimodal function with fixed dimensions proves that except for $f15(x)$ and $f18(x)$, in which PSO, and $f12(x)$ and $f13(x)$, in which GSA obtained better results, KGMO managed within 150 iterations to outperform PSO and GSA in 500 iterations, for the third type of benchmark functions. KGMO was at least 300% faster in terms of required running iterations to obtain the global minimum.

8.7 SUMMARY

A new optimization algorithm based on the natural behaviour of gas molecules under the kinetic energy is proposed. Performance evaluation of the proposed algorithm with other benchmark optimization algorithms using 23 benchmark functions, shows that KGMO finds the global minimum quickly, requiring less than 150 iterations, and able to decrease the MSE by 10^1 and 10^4 times as compared to PSO and GSA, respectively. In addition, KGMO is at least 600% faster for the unimodal and multimodal high dimensional functions, and at least 300% faster for the multimodal functions with fixed dimensions. In terms of mathematical equations, the efficiency of KGMO is also evident in terms of simplicity. Therefore, the proposed algorithm has better ability in finding the global optima compared to GSA and PSO without being trapped in local minima.

REFERENCES

Blondin, J. (2009). *Particle Swarm Optimization: A Tutorial*. Amsterdam University.

Bonabeau, E., Dorigo, M., & Theraulaz, G. (1999). *Swarm Intelligence: From Natural to Artificial Systems*. Oxford University Press.

Chun-an, L. (2008). New dynamic constrained optimization PSO algorithm. *Proc. Fourth International Conference on Natural Computation*, 650-653.

Cong, L., Sha, Y., & Jiao, L. (2006). Numerical optimization using organisational particle swarm algorithm. Simulated Evolution and Training. *Lecture Notes in Computer Science*, (4247), 150–157. doi:10.1007/11903697_20

Corne, D., Dorigo, M., & Glover, F. (Eds.). (1999). New Ideas in Optimization. McGraw Hill.

Devi, S., Jadhav, D. G., & Pattnaik, S. S. (2011). PSO based memetic algorithm for unimodal and multimodal function optimization. Swarm, Evolutionary, And Memetic Computing. *Lecture Notes in Computer Science*, *7076*, 127–134. doi:10.1007/978-3-642-27172-4_16

Dong, H. K., Abraham, A., & Jae, H. C. (2007). A hybrid genetic algorithm and bacterial foraging approach for global optimization. *Information Sciences*, (177): 3918–3937.

Dorigo, M., Maniezzo, V., & Colorni, A. (1996). The ant system: Optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man, and Cybernetics*, *26*(1), 29–41. doi:10.1109/3477.484436 PMID:18263004

Eberhart, R., & Shi, Y. (2001). Particle swarm optimization: developments, application and resources. *Proceedings of IEEE Congress of Evolutionary Computation*, 81-86. doi:10.1109/CEC.2001.934374

Eberhart, R., Shi, Y., & Kennedy, J. (2001). *Swarm Intelligence.* Academic Press.

Hopwood, S. J., & Jeans, J. (2009). *An Introduction to the Kinetic Theory of Gases*. Cambridge University Press.

Hu, G. M., Palreddy, S., & Tompkins, W. (1997). Patient adaptable ECG beat classifier using a mixture of experts approach. *IEEE Transactions on Biomedical Engineering*, (44): 891–900. PMID:9282481

Idoumghar, L., Melkemi, M., Schott, R., & Aouad, M. I. (2011). *Hybrid PSO-SA type algorithms for multimodal function optimization and reducing energy consumption in embedded systems. Applied Computational Intelligence and Soft Computing*. Hindawi.

Rashedi, E., Nezamabadi-pour, H., & Saryazdi, S. (2009). GSA: A gravitational search algorithm. *Information Sciences*, *179*(13), 2232–2248. doi:10.1016/j.ins.2009.03.004

Robert, M., & Ian, R. G. (2006). The effects of noise on computerised electrocardiogram measurements. *Journal of Electrocardiology*, *39*(4), 165–173. doi:10.1016/j.jelectrocard.2006.05.025

Salerno, J. (1997). Using the particle swarm optimization technique to train a recurrent neural network. *Proceedings of ninth IEEE international conference on tools with artificial intelligence*, 45-49. doi:10.1109/TAI.1997.632235

Chapter 9
Classification and Feature Extraction

ABSTRACT

In this chapter, the proposed optimization algorithm, kinetic gas molecule optimization (KGMO), that is based on swarm behaviour of gas molecules is applied to train a feedforward neural network for classification of ECG signals. Five types of ECG signals are used in this work including normal, supraventricular, brunch bundle block, anterior myocardial infarction (Anterior MI), and interior myocardial infarction (Interior MI). The classification performance of the proposed KGMO neural network (KGMONN) was evaluated on the Physiobank database and compared against conventional algorithms.

9.1 INTRODUCTION

In previous chapters, it was described that the shape of ECG waveform and heart rate reflects the state of the heart. It may contain important pointers to various types of disorders afflicting the heart. Bio-signals are non-stationary signals; therefore, the reflection may occur randomly in the time-scale. The disease symptoms may not show up all the time, but would manifest at certain irregular intervals during the day. Therefore, for effective diagnostics, ECG pattern and heart rate variability may have to be observed over several hours.

From the literature, it is evident that there is a need for efficient classification of ECG signals for effective and robust heart disorder recognition. This chapter

DOI: 10.4018/978-1-5225-5580-3.ch009

aims to use the efficient Kinetic Gas Molecule Optimization (KGMO), to train a neural network in the classification of ECG signals for the automated identification of heart disorders. The proposed intelligent system is called KGMO neural network (KGMONN).

For classification, it is necessary that useful features of all ECG signals be extracted. The feature extraction module is concerned with forming a vector of measurements (feature vector) from each heartbeat, that is processed by the classifier. The feature extraction module is required because, although it is possible for the classification stage to process the ECG samples directly, greater classification performance is often achieved if a smaller number of discriminating features are first extracted from the ECG samples.

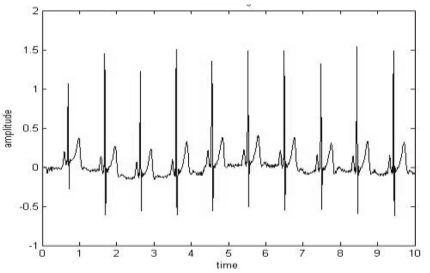
9.2 FEATURES OF DISORDERS

Consultation with heart specialists determined that 4 types of heart disorders, namely, bundle branch block, supraventricular tachycardia, anterior myocardial infarction (anterior MI) and inferior myocardial infarction (inferior MI) are common and would be detected in this work. Sample ECG signals for the 4 classes and normal ECG signals are given in Figure 1. Specialists detect these disorders by observing the PQRST waveform. For example, bundle branch block causes a widened and possibly jagged QRS waveform, while supraventricular tachycardia typically exhibits a narrow QRS complex on the ECG. The ST segment, which is normally iso-electric (flat and in line with the PQ segment) may be elevated or depressed due to myocardial ischemia or myocardial infarction. Table 1 shows the ECG classes and representation of the desired neural network outputs for each class.

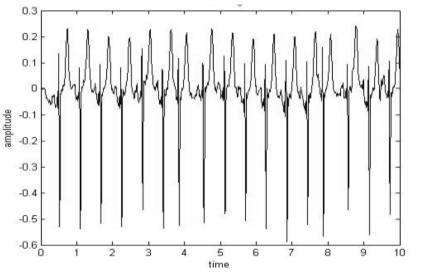
The feature selection is very important step in clustering process. The aim is to select the only most important features to reduce their number and at the same time, retain as much of their class discriminatory information as possible. If features with little discrimination power are selected, the subsequent cluster design would lead to poor performance.

The extracted features in this work are based on consultation with a heart specialist. The selected features are QRS duration, PR duration, QT duration, heart rate, RR duration and ST duration (Positive/Negative T; i.e. if the T is negative, then ST duration is 0). Appendix B shows the dataset of the 4 classes of heart disorders and the normal class. Figures 2-6 present each feature for

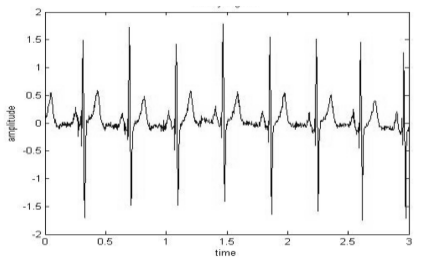
Figure 1. Typical waveform of ECG signals



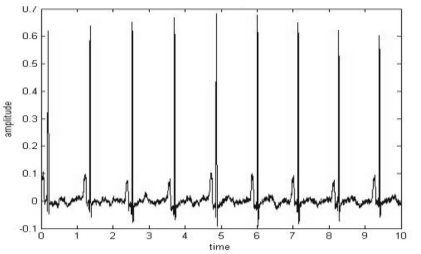
(a) Normal sinus



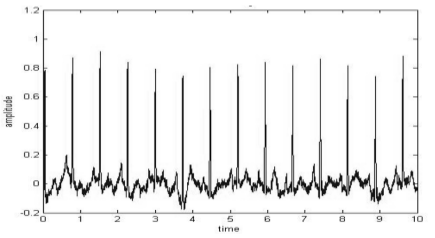
(b) Bundle branch block



(c) Supraventricular tachycardia



(d) Inferior myocardial infarction

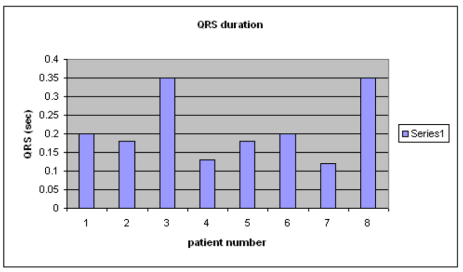


(e) Anterior myocardial infarction

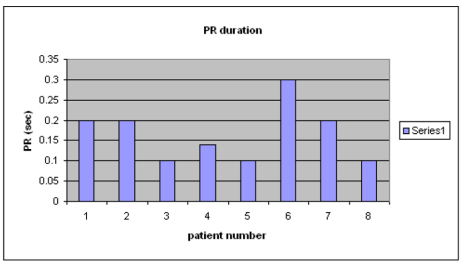
Table 1. ECG classes and representation of desired neural network outputs

Disorder	Class	Neural Network Output		
Normal sinus	0	0	0	0
Bundle branch block	1	0	0	1
Supraventricular tachycardia	2	0	1	0
Anterior myocardial infarction	3	0	1	1
Inferior myocardial infarction	4	1	0	0

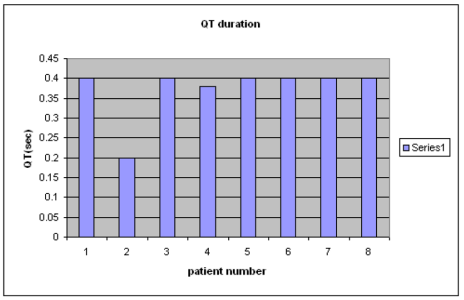
Figure 2. Selected features for brunch bundle block



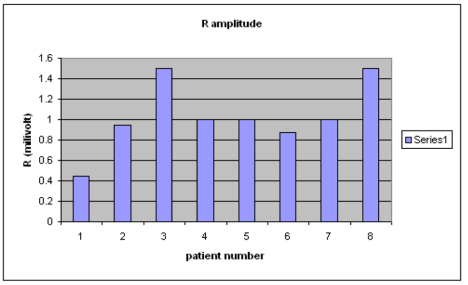
(a) QRS duration



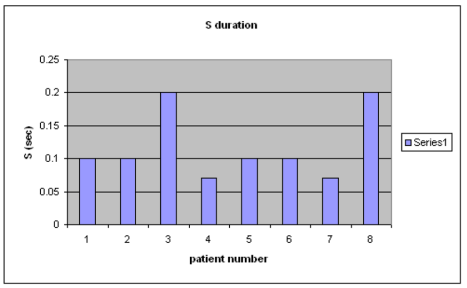
(b) PR duration



(c) QT duration



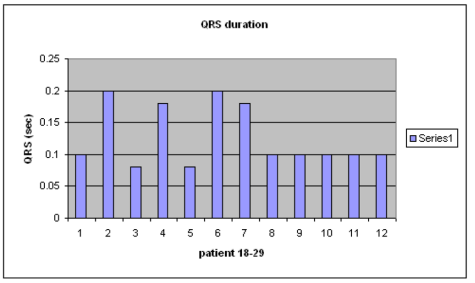
(d) R amplitude



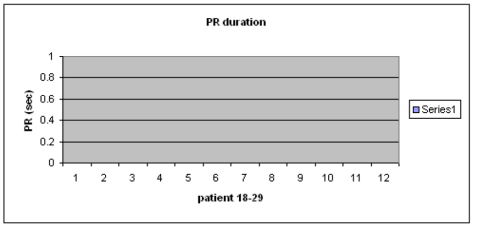
(e) ST duration

a number of samples affected by each of mentioned disorders, including the normal human heart beat. In Figure 2, it is obvious that for Bundle branch block, the QRS duration is wider than 0.12 seconds. Figure 3 shows lack of PR duration in supraventricular tachycardia. In Figure 4, it is clear that inferior MI causes a narrower QRS than normal as well as lack of normal ST duration (because of negative T). Figure 5 shows the samples of anterior MI that have no QT segment. Finally, Figure 6 shows a normal human heart beat, where the features are in the normal range (as explained in chapter 2).

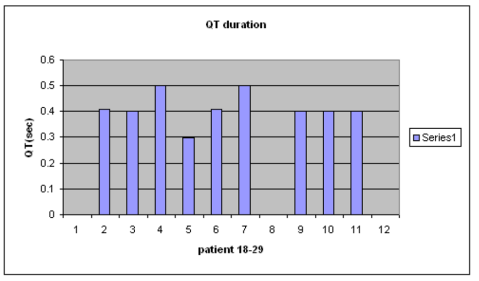
Figure 3. Selected features for supraventricular tachycardia



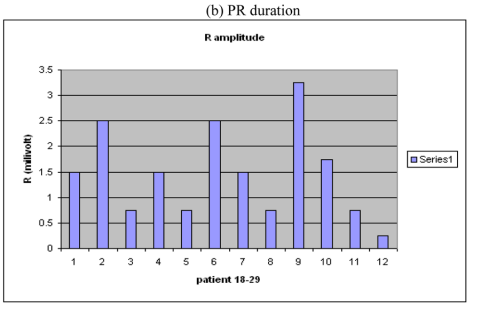
(a) QRS duration



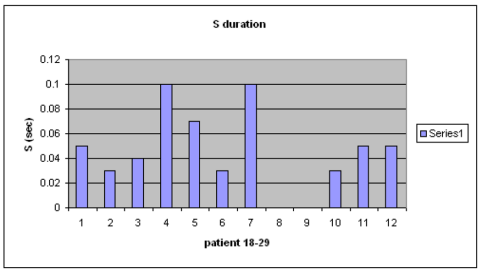
(b) PR duration



(c) QT duration



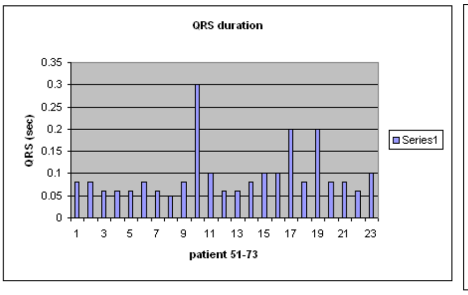
(d) R amplitude



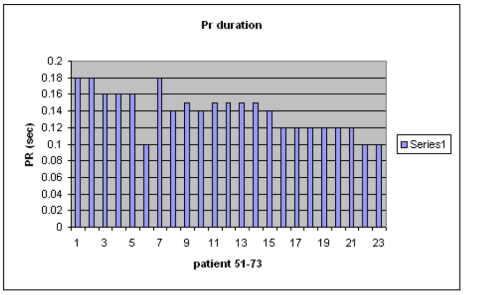
(e) ST duration

Another presentation of the extracted features for disorders is provided in Figuure 7. Figure 7(a) shows the QRS duration for various samples. It is obvious that that patient 20-35 has the wider QRS, which is the symptom of bundle branch block. Figure 7 (b) shows the PR duration for various patients. It presents that patient 35-50 has the very narrow or zero PR, which is the symptom of supraventricular tachycardia. Figure 7 (c) and (d) are QT interval and T wave presentations, respectively, that are used for detecting inferior and anterior MI. Figures show the negative T and longer QT interval for patient 50-75 which is the symptoms of inferior MI, and presents the negative T with lack of QT interval (or small QT) for patients 75-100.

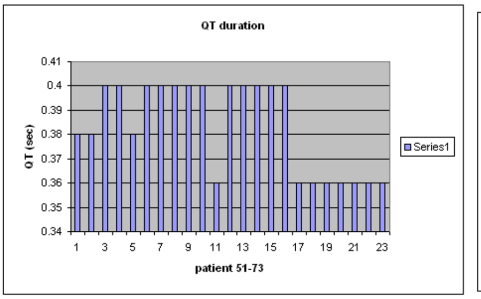
Figure 4. Selected features for inferior MI



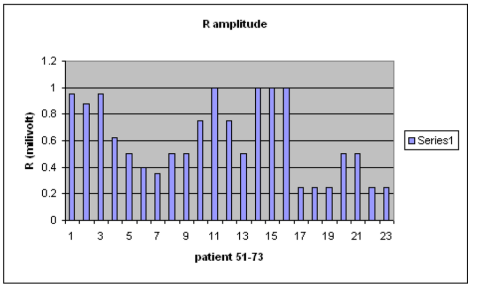
(a) QRS duration



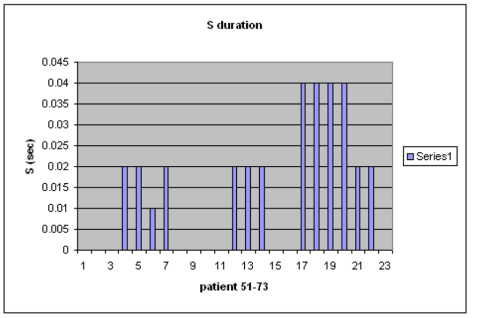
(b) PR duration



(c) QT duration



(d) R amplitude

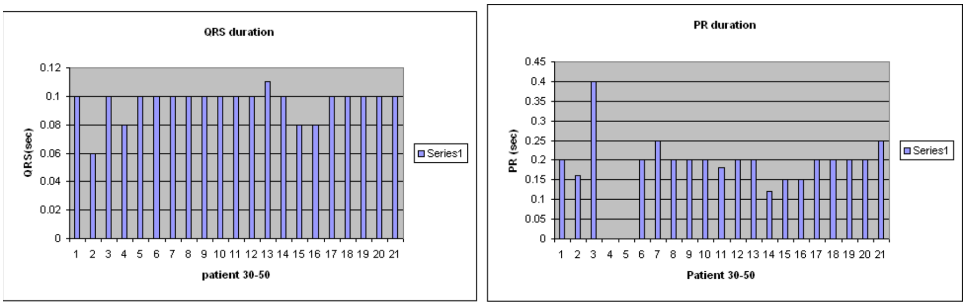


(e) ST duration

9.3 ECG SIGNAL CLASSIFICATION

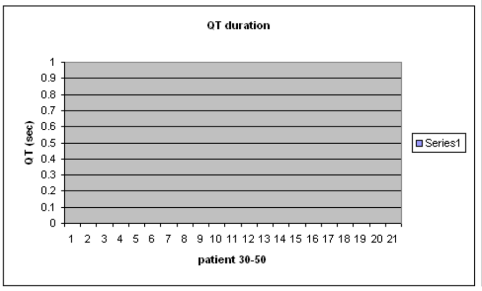
Heart diseases/abnormalities should be detected early for the prolonging of life and enhancing the quality of living through appropriate treatment. For effective diagnosis, the study of the ECG pattern and heart rate variability signals may have to be carried out over several hours (Tomas and Neil, 2002). The accurate classification of ECG signals is significant for decision making.

Figure 5. Selected features for anterior MI

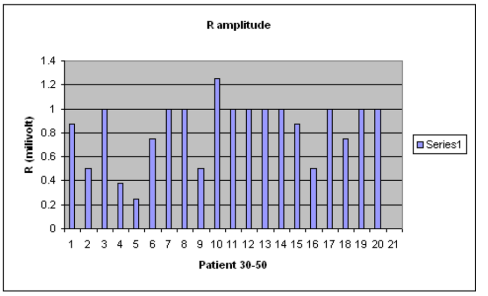


(a) QRS duration

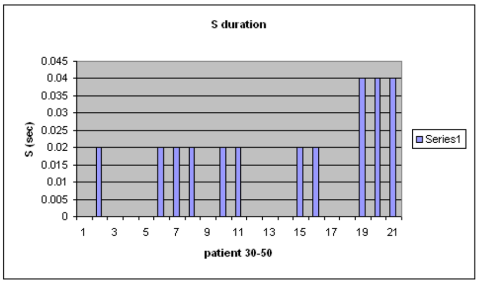
(b) PR duration



(c) QT duration



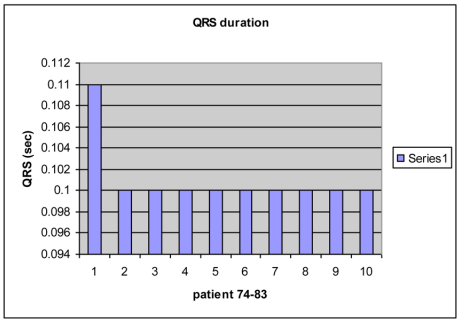
(d) R amplitude



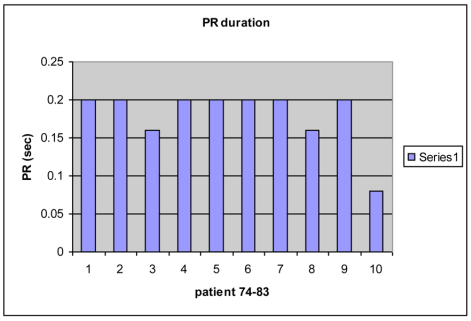
(e) ST duration

A total of 100 ECG signals were randomly selected from the various classes of databases for testing the performance of SOM in identifying ω_c using the proposed method. The outcome of scanning the clustered frequency spectra are normalised as described in the chapter 3. Once the cutoff frequency has been determined, the test signals are denoised using the FIR filter. A Graphical User Interface (GUI) is designed to use the SOM method for the noise removal process and the 6 mentioned features (the duration of QRS, PR, QT, RR, heart rate and T absence) are extracted. The features are collected in a dataset, as in Appendix B. The last column shows the disorder related to the signal.

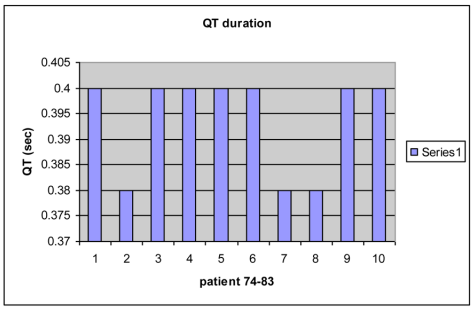
Figure 6. Selected features for normal sinus



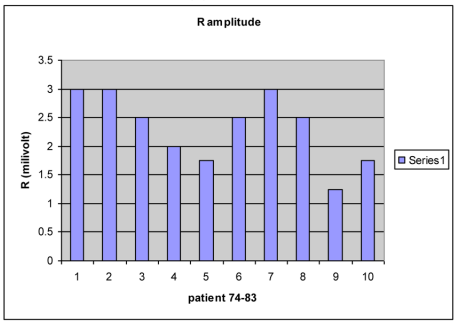
(a) QRS duration



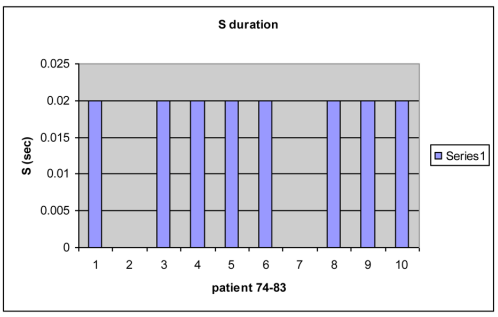
(b) PR duration



(c) QT duration



(d) R amplitude

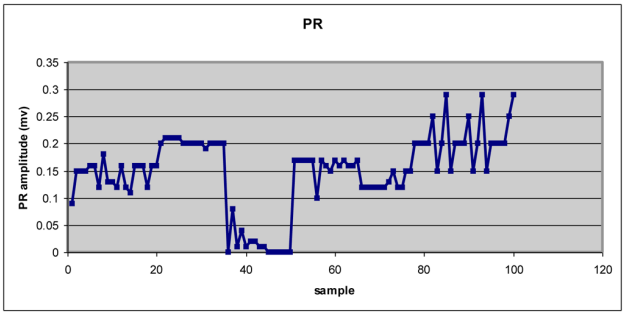


(e) ST duration

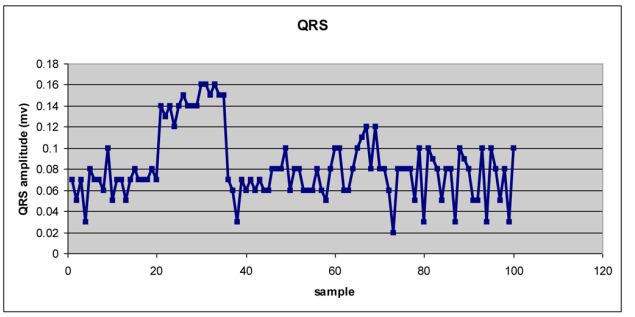
9.3.1 Feedforward Neural Network Training by KGMO

For a neural network implementation, each gas molecule in the swarm has a position that represents a set of weights for the current training cycle. The number of weights associated with the network defines the dimensionality of each gas molecule. The gas molecule aims to reduce the learning error (mean squared error, MSE) and for this, it moves within the weight space.

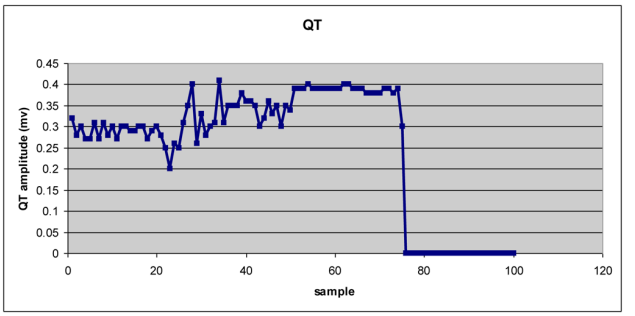
Figure 7.



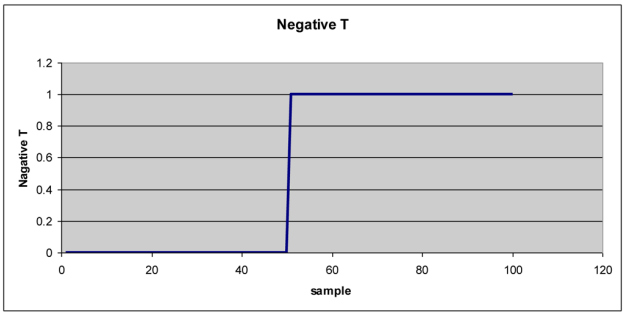
(a) PR interval duration for 100 patients



(b) QRS duration for 100 patients



(c) QT interval duration for 100 patients



(d) T wave for 100 patients (T=1 shows the negative T wave)

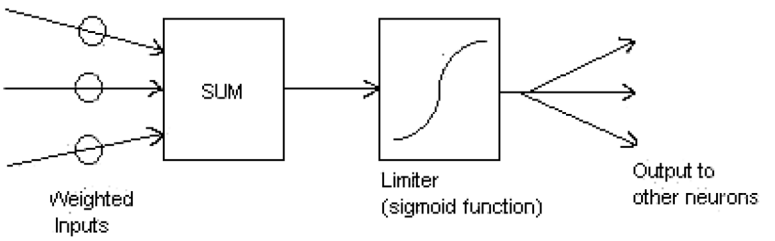
By changing the position in each training cycle, the weights of the network are updated in order to reduce the error. In each training cycle, all the gas molecules update their position by calculating the new kinetic energy of each gas molecule and then calculate the velocity of the move to the new position. The new position is a set of new weights used to obtain the new error value. For KGMO, the new weights are adapted even though no improvement is observed (e.g. when trapped in local optima). This process is repeated for all the gas molecules. The global best gas molecule is the one with the lowest error thus far. The training process stops when the error falls below a set threshold or computational limits are exceeded. When the training ends, the weights are used to calculate the classification error for the training patterns. The same set of weights is then used to test the network with the test patterns.

9.3.2 KGMONN for ECG Signal Classification

The KGMONN uses a multi-layer feedforward network architecture, beginning with an input layer. This layer is connected to a hidden layer. This hidden layer can then be connected to another hidden layer or directly to the output layer. There can be any number of hidden layers so long as at least one hidden layer is provided. In common use, most neural networks will have only one hidden layer. It is very rare for a neural network to have more than two hidden layers (Hussoun, 1995). Figure 8 shows the mathematical structure of a neuron in a feedforward neural network.

NN produces the learning error (gas molecule fitness) based on a set of weights and bias (KGMO positions). Each gas molecule's lowest learning error so far (*pbest*) and the lowest learning error found in the entire learning process so far (*gbest*) are applied to the kinetic energy update Equation (12) in chapter 8 and then to the position update Equation (13) in chapter 8 to update the velocity Equation (14) in chapter 8 in achieving the best solution

Figure 8. Mathematical form of a neuron in a feedforward neural network



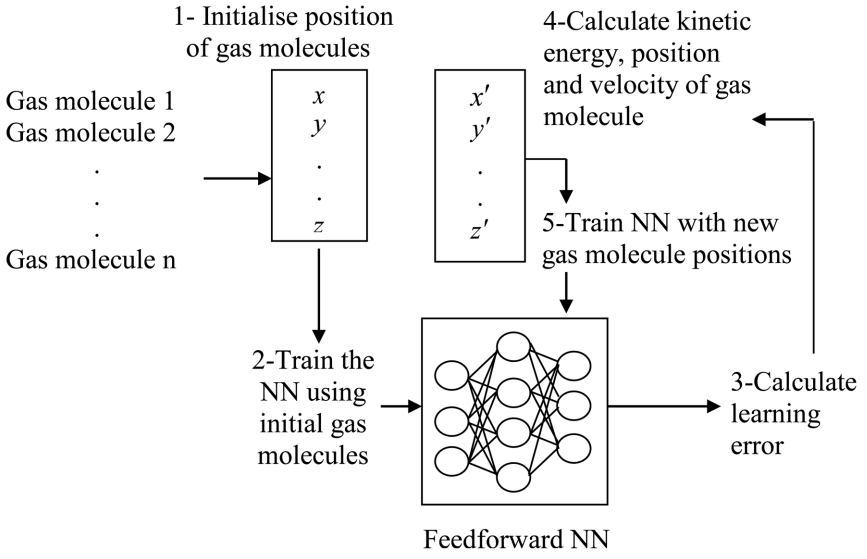
or targeted learning error. The new sets of positions (weights and biases) are produced by adding the calculated velocity value from Equation (14) in chapter 8 to the current position value using the movement Equation (13) in chapter 8. Then, the new set of positions is used to produce new learning error (gas molecule fitness) in the feedforward neural network (NN). This process is repeated until the stop condition is met (minimum learning error or maximum number of iteration). The summary of the KGMONN learning process is given by Figure 9.

9.3.3 KGMONN Configuration

As 6 features are extracted from each ECG signals, the feedforward NN consists of 6 nodes in the input layer. For the 5 classes of heart disorders, 3 nodes are used in the output layer for the binary output as it is presented in Table 3. The NN is trained by the KGMO algorithm (based on chapter 4 initialisation for KGMO) with $c_1 = 1$, $c_2 = 3$, $m = 1$, inertia factor (w) decreased linearly from 0.85 to 0.2 and E decreased linearly from 0.95 to 0.1.

The training and testing datasets are from the 100 vectors. Due to the lack of certain cases in the configured datasets, validity testing of the results was undertaken using a k-folding method with k=5. For each of the k times of NN

Figure 9. Feedforward neural network training (NN) by KGMO



training, 80% of the samples in the dataset were used for training with the non-overlapping 20% for testing. All the results were obtained by averaging the outcomes of five separate tests. Mean squared error (MSE) is used as the measure of the error with a target MSE of 0.01 or less, given by (1).

$$MSE = \frac{1}{N}\sum_{n=0}^{N-1}(y_n - t_n)^2 \tag{1}$$

where y_n is the output of the network, t_n is the desired target, and N is the total number of test records.

9.3.4 Benchmark Algorithms

To evaluate the relative performance of the KGMONN, it was compared against NN trained with the popular PSO and back propagation (BP) algorithms. BP is the widely used supervised learning algorithm (Hassoun, 1995) while PSO is a population based stochastic optimization technique as explained in previous chapter. It is also explained that PSO updates the population of particles by updating the following equations:

$$x_i^d(t+1) = x_i^d(t) + v_i^d(t+1) \tag{2}$$

$$v_i^d(t+1) = wv_i^d(t) + c_1 rand_i(pbest_i^d - x_i^d(t)) + c_2 rand_i(gbest^d - x_i^d(t)) \tag{3}$$

The population size used was 25, with the default value of 2 for the maximum velocity divisor, c_1 and c_2. The PSO algorithm is used for training the feedforward NN with the same concepts as explained in other sections of this chapter.

9.4 RESULTS

The classification results achieved for the ECG signals using KGMONN, PSONN and BP in six different topologies are presented in Table 2, Table 3 and Table 4, respectively. The *max training cycle* parameter indicates the maximum number of training cycles for stopping the training process. The

Table 2 Average training results for six topologies of KGMONN over 10 times running

NI	NH	NO	Max training cycle	Learning training cycle	MSE (train)	MSE (test)	Correctly classified (train) (%)	Correctly classified (test) (%)
6	6	3	300	300	0.1521	0.1540	0.8563	0.8512
6	5	3	300	300	0.1442	0.1455	0.8674	0.8622
6	4	3	300	300	0.1405	0.1428	0.8682	0.8652
6	6	3	1000	550	0.1541	0.1562	0.8548	0.8539
6	5	3	1000	550	0.1585	0.1612	0.8563	0.8456
6	**4**	**3**	**1000**	**550**	**0.1272**	**0.1346**	**0.8696**	**0.8645**
6	3	3	1000	400	0.1675	0.1672	0.8411	0.8442
6	2	3	1000	400	0.1761	0.1762	0.8319	0.8318

NI: *Number of nodes in input layer,*
NH: *Number of nodes in hidden layer,*
NO: *Number of nodes in output layer.*

Table 3. Average training results for six topologies of PSONN over 10 runs

NI	NH	NO	Max training cycle	Learning training cycle	MSE (train)	MSE (test)	Correctly classified (train) (%)	Correctly classified (test) (%)
6	6	3	300	300	1.3804	1.4804	0.4011	0.3945
6	5	3	300	300	1.4891	1.5303	0.4453	0.2314
6	4	3	300	300	1.5482	1.5553	0.3549	0.3265
6	6	3	600	600	0.3221	0.3421	0.7044	0.6977
6	5	3	600	600	0.3071	0.3387	0.7198	0.7070
6	**4**	**3**	**600**	**600**	**0.2901**	**0.3021**	**0.7360**	**0.7299**
6	3	3	600	600	0.3312	0.3212	0.6884	0.6824
6	2	3	600	600	0.3011	0.3065	0.7028	0.7032

Table 4. Average training results of six topologies of NN trained using BP over 10 runs

NI	NH	NO	Max training cycle	Learning training cycle	MSE (train)	MSE (test)	Correctly classified (train) (%)	Correctly classified (test) (%)
6	6	3	1000	200	3.9693	3.9187	0.2777	0.2574
6	**5**	**3**	**1000**	**200**	**3.6298**	**3.3275**	**0.2936**	**0.2394**
6	4	3	1000	200	3.9375	3.9500	0.2637	0.2964

learning training cycles parameter indicates the number of training cycles that is passed for reaching to the minimum error of the network.

From the tables, the best topologies in terms of *{number of input nodes, number of hidden nodes, number of output nodes}* for KGMONN, PSONN and BP are *{6, 4, 3}, {6, 4, 3}* and *{6, 5, 3},* respectively, as the network error is minimum at these configurations. The MSE for training and testing are calculated by averaging over 10 runs. The *correctly classified* measure shows the correlation that is the relationship between two sets of data (the calculated output and the desired output). Comparing the results shows that KGMONN produced less error than PSONN and BP. The first 3 rows of Table 3 present test errors of PSONN over 300 training cycles of more than 1, much greater than the desired target error of 0.01. The KGMONN achieved a significant preference to PSONN and BP is related to its fast convergence property. Using the *{6, 4, 3}, {6, 4, 3}, {6, 5, 3}* topologies for KGMONN, PSONN and BP, the cost function value (MSE) for both networks over 600 training cycles are presented in Figure 10, Figure 11, and Figure 12, respectively. Comparing the figures proves that KGMONN is less costly than PSONN and BP, in addition to being able to converge much faster than PSONN (in about 50 training cycles, as opposed to over 400 training cycles for PSONN).

The classification results obtained can be represented by a confusion matrix, where each cell contains the raw number of exemplars classified for the corresponding combination of desired and actual network outputs. A confusion matrix (Kohavi and Provost, 1998) contains information about actual and predicted classifications by a classification system. Performance

Figure 10. MSE of ECG signal classification by KGMONN

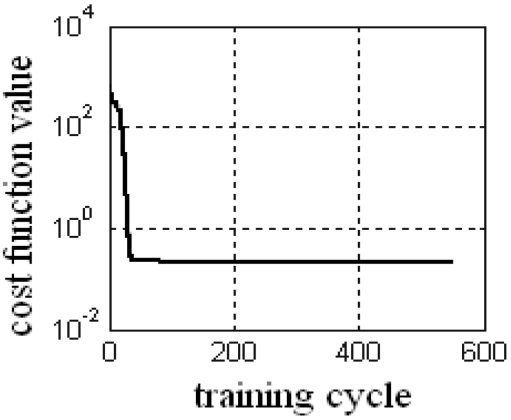


Figure 11. MSE of ECG signal classification by PSONN

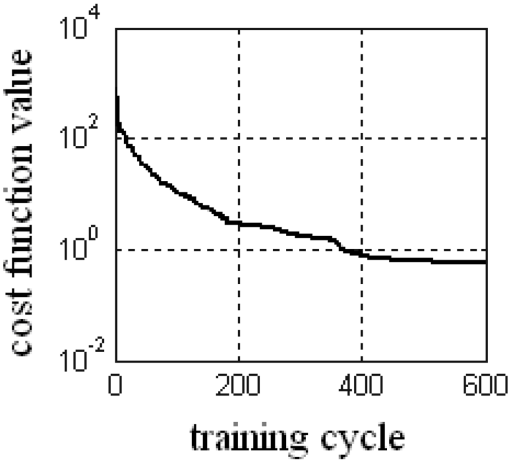
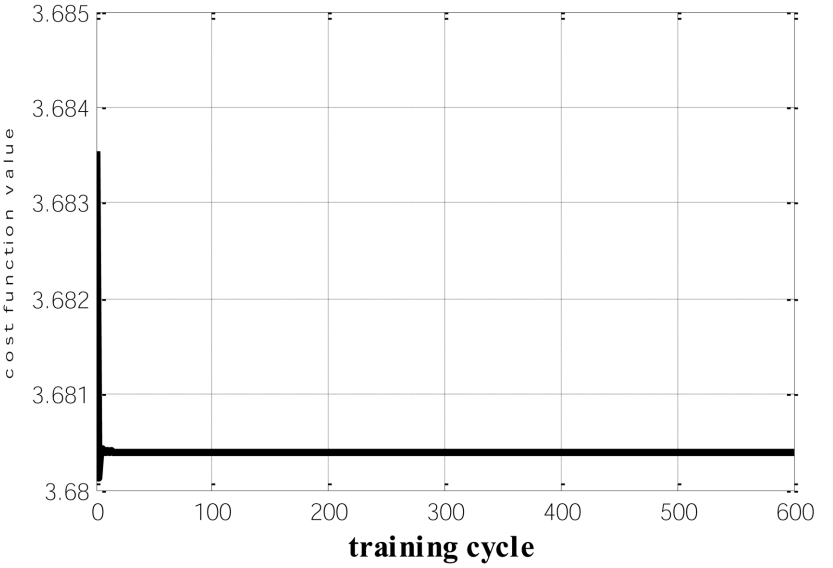


Figure 12. MSE of ECG signal classification by BP



of such systems is commonly evaluated using the data in the matrix. The rows present the number of actual classifications in the test data, while the columns present the number of predicted classifications made by the model. Table 5 shows the confusion matrix for the classification results using the KGMONN. Using the table, it is observed that the KGMONN can detect 17,

Table 5. The confusion matrix of classification by KGMONN

Desired results	Output results				
	Normal	Bundle branch block	Supraventricular	Inferior MI	Anterior MI
Normal	17	0	0	0	1
Bundle branch block	0	12	1	1	0
Supraventricular tachycardia	0	0	13	0	0
Inferior MI	1	1	0	22	3
Anterior MI	2	2	1	2	21

12, 13, 22 and 21 of signals from each class (Normal, BBB, SVT, Inferior MI and Anterior MI, respectively), correctly.

For further evaluation, the performance of the classifiers is assessed by the computation of specificity, sensitivity and total classification accuracy, defined as follows:

- *Sensitivity: (number of true positive)/(number of true positive+ number of false negatives)*
- *Specificity: (number of true negative)/(number of true negatives+ number of false positives)*
- *Total classification accuracy=(sensitivity + specificity)/2*

A true negative decision occurs when both the classifier and the physician suggested the absence of a positive detection. A true positive decision occurs when the positive detection of the classifier tallies with a positive detection by the physician. Table 6 shows the specificity, sensitivity and total accuracy of the classifier. The KGMONN achieved a total accuracy equal to 85%.

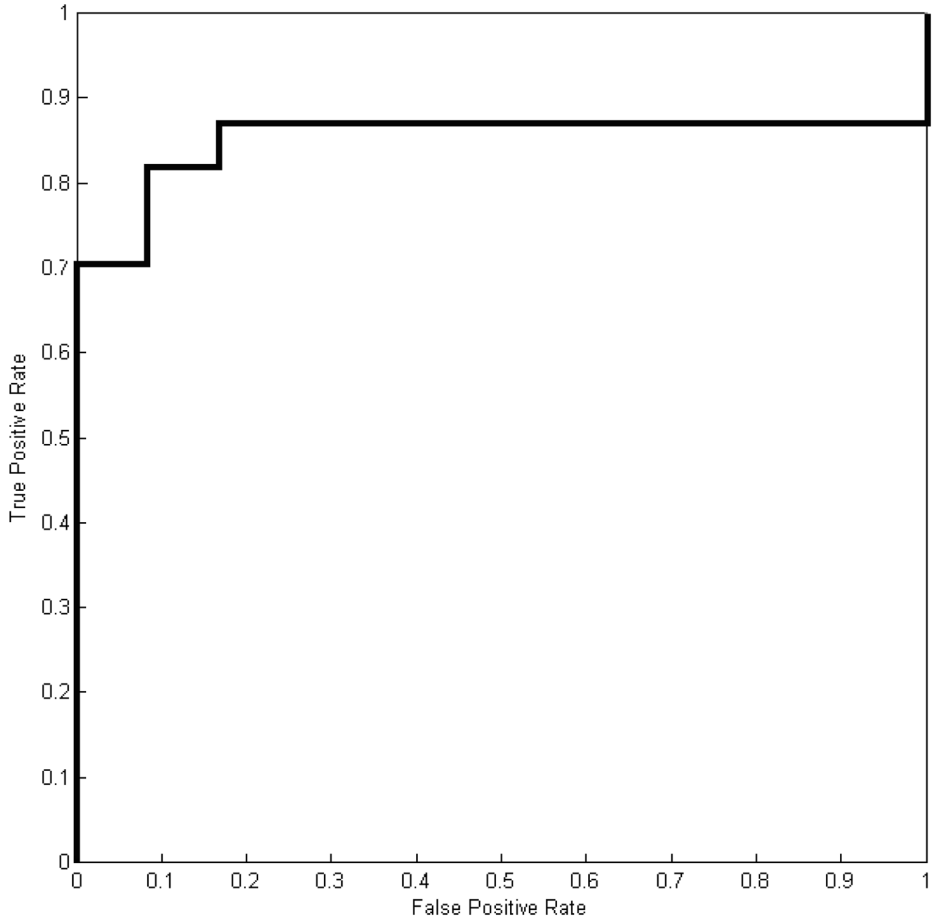
The performance of the classifier can be further shown via the spectrum of sensitivities and specificities plotted by the receiver operating characteristic (ROC) curve (Elif, 2010). ROC curves provide a comprehensive and visually attractive way to summarise the accuracy of predictions. They are widely applicable, regardless of the source of predictions. The closer the curve comes to the 45-degree diagonal of the ROC space, the less accurate the test.

A good test is when in the ROC curve, the sensitivity rises rapidly and the specificity hardly increases until the sensitivity becomes high. Figure 13 shows

Table 6. The values of statistical parameters

Desired results	Statistical Parameters		
	Sensitivity	Specificity	Total classification accuracy
Normal	85.00	85.00	84.82
Bundle branch block	80.00	85.88	
Supraventricular tachycardia	86.66	84.70	
Inferior MI	88.00	84.00	
Anterior MI	84.00	85.33	

Figure 13. ROC curve for classifier performance



the ROC curve for evaluating the performance of the KGMONN classifier. Since the true positive rate rises more rapidly than the false positive rate, it is proven that the proposed classifier has good accuracy for classification of the ECG signals.

9.5 SUMMARY

In this chapter, the KGMO is used to train the feedforward NN for classification of 5 types of ECG signals. Comparison of the results presents that the proposed KGMONN is a fast converging algorithm that produces less error than the conventional PSONN and BP algorithms. The statistical parameters show that the proposed classifier has considerable success in classification of ECG signals. The proposed training algorithm may be used in future for different architectures of NN and for a variety of applications.

REFERENCES

Castro, B., Kogan, D., & Geva, A. B. (2000). ECG feature extraction using optimal mother wavelet. *The 21st IEEE Convention of the Electrical and Electronic Engineers*, 346-350. 10.1109/EEEI.2000.924422

Liu, H., & Motoda, H. (1998). *Feature Extraction, Construction and Selection: A Data Mining Perspective*. Kluwer Academic. doi:10.1007/978-1-4615-5725-8

Mahmoodabadi, S. Z., Ahmadian, A., & Abolhasani, M. D. (2005). ECG feature extraction using daubechies wavelets. *Proceedings of the fifth IASTED International conference on Visualization, Imaging and Image Processing*, 343-348.

Manolescu, D. A. (n.d.). *Feature Extraction—A Pattern for Information Retrieval*. Academic Press.

Tadejko, P., & Rakowski, W. (2007). Mathematical morphology based ECG feature extraction for the purpose of heartbeat classification. *6th International Conference on Computer Information Systems and Industrial Management Applications*, 322-327. 10.1109/CISIM.2007.47

Zhao, Q., & Zhan, L. (2005). ECG feature extraction and classification using wavelet transform and support vector machines. *International Conference on Neural Networks and Brain*, (2), 1089-1092.

Conclusion

1. INTRODUCTION

This conclusion summarises the work accomplished in this research. The main contributions are summarised in section 2. Recommendations for future work are given in section 3. Finally, the overall conclusion of the research is provided at the end of this chapter.

2. SUMMARY OF CONTRIBUTIONS

The main contribution of this work is on automated noise removal and intelligent computer-aided classification of ECG signals using neural network. A new swarm based algorithm based on the behaviour of gas molecules was designed and a corresponded neural network was implemented. The proposed algorithms and methodology was proven to perform better on benchmark tests. The specific contributions are summarised as follows.

2.1 ECG Signal Noise Removal

Two intelligent methods for identifying the cutoff frequency are proposed for low-pass filtering using the FIR filter. The methods are as follows:

1. **SOM for finding cutoff frequency:** A self-organised map (SOM) neural network was set up to automatically identify the cutoff frequency to be used by an FIR filter in denoising the ECG signals. The results present that the identified cutoff frequency is accurate and better than the conventional method as the SOM-based approach works on each signal individually. Also, no loss of data is encountered in the denoised signal during the ECG low-pass filtering.

2. **PSONN for finding the cutoff frequency:** In this research it is shown that the Particle Swarm Optimization Neural Network (PSONN) is a reliable and fast solution for automatically identifying the cutoff frequency. The results show that the PSONN is better able to remove the noise than the conventional method. It is also evident that the proposed method is promising for use by medical experts who wish to diagnose heart disorders using ECG.

Comparing the MSE of SOM and PSONN for finding the cutoff frequency shows that PSONN has the smaller MSE (around 0.16) comparing with SOM (around 0.4). However, because of the complexity of PSONN in implementation and more required time for finding the cutoff frequency, SOM is preferred for denoising the ECG signals in the proposed classification system.

2.2 ECG Signal Feature Extraction

Identification of the best features for classification of heart disorders was undertaken via consultation with heart specialist. As the result of feature extraction and collecting the suitable data from Physiobank database, an appropriate dataset is prepared. The ECG signal features identified for the heart disorders were QRS duration, QT duration, PR duration, RR duration, heart rate and presence or absence of T wave.

2.3 KGMO

The main contribution in this work is development of a new optimization algorithm based on the natural behaviour of gas molecules acted on by kinetic energy. The performance of the proposed Kinetic Gas Molecule Optimization (KGMO) is evaluated using 23 benchmark functions, and compared against two other benchmark swarm based algorithms. For the unimodal benchmark functions, KGMO finds the best global minimum in the less than 150 iterations with better results compared to Particle Swarm Optimization (PSO) and Gravitational Search Algorithm (GSA). For multimodal high dimensional functions, again KGMO has better performance. For the third group of benchmark functions, i.e. multi-modal with fixed dimensions functions, KGMO finds better global minima for some functions, while PSO and GSA

has better performance for other functions. The overall evaluation of results using the Wilcoxon test and the Mean Square Error (MSE) shows that KGMO has better performance than PSO and GSA. It shows that KGMO finds the global minimum quickly, requiring less than 150 iterations, and able to decrease the MSE by 10^1 and 10^4 times as compared to PSO and GSA, respectively. The smaller standard deviation obtained by the KGMO compared to the other algorithms, especially in the first and second group of benchmark functions, proves the robustness and the strength of KGMO in quick and effective convergence.

In terms of mathematical equations and complexity, the efficiency of KGMO is also evident in terms of simplicity. Therefore, the proposed algorithm has better ability in finding the global optima compared to GSA and PSO without getting trapped in local minima.

2.4 KGMONN for Classification of ECG Signals

The proposed KGMO is used to train a feedforward NN for classification of 5 types of ECG signals. The introduced neural network (KGMONN) is a fast converging architecture as proven by the results. It also has less error compared with the traditional backpropagation neural network. The performance is statistically measured and shows that the proposed classifier has considerable success in classification of the ECG signals with the accuracy equal with 85%.

3. FUTURE WORKS AND LIMITATIONS

The research work undertaken has met all the set objectives. However, it could be further extended to further improve the performance and overcome some of the limitations. Some recommendations for future work are as follow.

- *Classification of ECG signal for larger number of disorders*.

The present research only classifies 5 types of ECG signals. A larger number of heart disorders could be identified, possibly with the inclusion of signals from the other leads as well. In this way, a larger dataset could be prepared for the classification of various disorders.

- *Automatic feature extraction.*

Could improve the current manual feature extraction ECG classification system, in the system, as implemented in the prototype, by developing an algorithm for automatic feature extraction. Thus, a fully automatic system could be implemented.

- *KGMO for noise removal.*

As the filtering stage also involves with optimization, it may be possible to apply KGMO in the filtering to minimise the difference between the noisy signal and the clean signal when noise removal is performed.

- *Recurrent neural network.*

Training by KGMO is suggested for classification of data. As it was seen, the present research used a feedforward neural network. It is expected that the performance of the classification system would improve using the suggested recurrent KGMONN.

- *Other applications of KGMO.*

In optimization could be tested. For example, the application of KGMO in networking, image processing, etc. could be investigated, for better performance.

4. FINAL WORDS

The main focus of this research was in developing a new optimization algorithm to train the neural network for ECG signal classification. Results prove the validity of the proposed KGMO algorithm and its ability for training a feedforward neural network for classification. Another success of the research includes using neural networks in noise removal. The results obtained are considered to be good as typically it is difficult to develop automated system for heart disorder detection using ECG signals due to inter and intra-patient variation in the signal readings. Some suggestions of future work for improving the performance of the findings and prototype are discussed.

APPENDIX A

Comparison of GA and PSO

Table 1. Comparison of GA and PSO

Algorithm	Strength	Weaknesses
PSO	According to Bai (2010) and Yu et al. (2008), Ismail and Shamsuddin (2008): a) It is easy to implement. b) It sets a few parameters (c_1, c_2 and w). c) It has been successfully applied in many areas: function optimization, ANN training, fuzzy system control d) It has no overlapping and mutation calculation and the search can be carried out by the speed of the particle. e) During the development of several generations, only the most optimist particle can transmit information onto the other particles. Therefore, it uses a little memory. f) The speed of the researching is very fast.	According to Bai (2010), Yu et al. (2008), Ismail and Shamsuddin (2008), Lian et al. (2008) and Cong et al. (2006): a) A potentially dangerous property in PSO is stagnation due to the lack of momentum, which makes it impossible to arrive at the global optimum. b) It has low convergence speed when it reaches, the optimum solution region. c) It cannot identify the nonlinear system with higher precision with another algorithm. d) The method easily suffers from partial optimism, which causes less exact at the regulation of its speed and the direction.
GA	According to *Sivanandam* and *Deepa* (2007), Yu et al. (2008): a) GA is intrinsically parallel and can explore the solution space in multiple directions at once. b) Solution space is wider. c) It is good for multi-modal problems and returns a suite of solution d) It handles noisy functions well. e) It is resistant to becoming trapped in local minima f) GA performs very well for large-scale optimization problems.	According to Mahmoudi et al. (2009) and *Sivanandam* and *Deepa* (2007): a) The GA may be unable to find a solution for the problem, or may end up solving the wrong problem. b) Definition of representation for the problem is hard. c) It needs a long training time to produce the output. d) It uses complex functions in selection, crossover and mutation calculation.

APPENDIX B

Dataset Prepared by Consultation With Heart Specialist

Table 2. The dataset of ECG patterns

QRS	PR	QT	rate	RR	Negative /Positive T	Disorder
0.07	0.09	0.32	54	1.09	0	Normal
0.05	0.15	0.28	62	0.96	0	Normal
0.07	0.15	0.30	97	0.61	0	Normal
0.03	0.15	0.27	97	0.61	0	Normal
0.08	0.16	0.27	95	0.62	0	Normal
0.07	0.16	0.31	79	0.75	0	Normal
0.07	0.12	0.27	71	0.83	0	Normal
0.06	0.18	0.31	81	0.73	0	Normal
0.10	0.13	0.28	68	0.87	0	Normal
0.05	0.13	0.30	88	0.67	0	Normal
0.07	0.12	0.27	78	0.80	0	Normal
0.07	0.16	0.30	70	0.70	0	Normal
0.05	0.12	0.30	77	0.69	0	Normal
0.07	0.11	0.29	70	0.80	0	Normal
0.08	0.16	0.29	80	0.61	0	Normal
0.07	0.16	0.30	81	0.70	0	Normal
0.07	0.16	0.30	70	0.70	0	Normal
0.07	0.12	0.27	72	0.80	0	Normal
0.08	0.16	0.29	77	0.60	0	Normal
0.07	0.16	0.30	73	0.70	0	Normal
0.10	0.20	0.28	78	0.76	0	Bundle branch block
0.10	0.21	0.25	78	0.76	0	Bundle branch block
0.10	0.21	0.20	78	0.71	0	Bundle branch block
0.10	0.21	0.26	72	0.75	0	Bundle branch block
0.12	0.21	0.25	67	0.76	0	Bundle branch block
0.10	0.20	0.31	69	1.09	0	Bundle branch block
0.10	0.20	0.35	56	1.06	0	Bundle branch block
0.10	0.20	0.40	65	0.91	0	Bundle branch block
0.10	0.20	0.26	78	0.76	0	Bundle branch block
0.10	0.20	0.33	51	1.16	0	Bundle branch block

continued on following page

Table 2. Continued

QRS	PR	QT	rate	RR	Negative /Positive T	Disorder
0.11	0.19	0.28	62	0.96	0	Bundle branch block
0.10	0.20	0.30	59	1.00	0	Bundle branch block
0.10	0.20	0.31	90	0.66	0	Bundle branch block
0.10	0.20	0.41	47	1.26	0	Bundle branch block
0.10	0.20	0.31	54	1.09	0	Bundle branch block
0.07	0.00	0.35	59	1.01	0	Supraventricular tachycardia
0.06	0.08	0.35	68	0.87	0	Supraventricular tachycardia
0.03	0.01	0.35	95	0.62	0	Supraventricular tachycardia
0.07	0.04	0.38	45	1.32	0	Supraventricular tachycardia
0.06	0.01	0.36	92	0.65	0	Supraventricular tachycardia
0.07	0.02	0.36	81	0.73	0	Supraventricular tachycardia
0.06	0.02	0.35	67	0.87	0	Supraventricular tachycardia
0.07	0.01	0.30	69	1.01	0	Supraventricular tachycardia
0.06	0.01	0.32	76	0.77	0	Supraventricular tachycardia
0.06	0.00	0.36	66	0.90	0	Supraventricular tachycardia
0.08	0.00	0.33	85	0.70	0	Supraventricular tachycardia
0.08	0.00	0.35	75	0.66	0	Supraventricular tachycardia
0.08	0.00	0.30	79	0.78	0	Supraventricular tachycardia
0.10	0.00	0.35	88	0.67	0	Supraventricular tachycardia
0.06	0.00	0.34	76	0.78	0	Supraventricular tachycardia
0.08	0.17	0.39	85	0.70	1	Inferior MI
0.08	0.17	0.39	85	0.70	1	Inferior MI
0.06	0.17	0.39	85	0.70	1	Inferior MI
0.06	0.17	0.40	85	0.70	1	Inferior MI
0.06	0.17	0.39	66	0.90	1	Inferior MI
0.08	0.10	0.39	76	0.78	1	Inferior MI
0.06	0.17	0.39	76	0.78	1	Inferior MI
0.05	0.16	0.39	86	0.70	1	Inferior MI
0.08	0.15	0.39	59	1.01	1	Inferior MI
0.31	0.17	0.39	76	0.78	1	Inferior MI
0.10	0.16	0.39	85	0.70	1	Inferior MI
0.06	0.17	0.40	76	0.78	1	Inferior MI
0.06	0.16	0.40	66	0.90	1	Inferior MI
0.08	0.16	0.39	66	0.90	1	Inferior MI
0.10	0.17	0.39	66	0.90	1	Inferior MI

continued on following page

Table 2. Continued

QRS	PR	QT	rate	RR	Negative /Positive T	Disorder
0.19	0.12	0.39	59	1.01	1	Inferior MI
0.20	0.12	0.38	85	0.70	1	Inferior MI
0.08	0.12	0.38	85	0.70	1	Inferior MI
0.20	0.12	0.38	85	0.70	1	Inferior MI
0.08	0.12	0.38	85	0.70	1	Inferior MI
0.08	0.12	0.39	85	0.70	1	Inferior MI
0.06	0.13	0.39	76	0.78	1	Inferior MI
0.92	0.15	0.38	59	1.01	1	Inferior MI
0.08	0.12	0.39	73	0.70	1	Inferior MI
0.08	0.12	0.30	80	0.70	1	Inferior MI
0.08	0.15	0.00	71	0.83	1	Anterior MI
0.08	0.15	0.00	88	0.67	1	Anterior MI
0.95	0.20	0.00	47	1.26	1	Anterior MI
0.10	0.20	0.00	85	0.70	1	Anterior MI
0.93	0.20	0.00	59	1.01	1	Anterior MI
0.10	0.20	0.00	47	1.26	1	Anterior MI
0.90	0.25	0.00	47	1.26	1	Anterior MI
0.08	0.15	0.00	77	0.60	1	Anterior MI
0.91	0.20	0.00	67	1.02	1	Anterior MI
0.98	0.29	0.00	54	1.20	1	Anterior MI
0.08	0.15	0.00	72	0.67	1	Anterior MI
0.08	0.2	0.00	72	0.67	1	Anterior MI
0.1	0.2	0.00	71	0.83	1	Anterior MI
0.09	0.2	0.00	85	0.7	1	Anterior MI
0.08	0.25	0.00	72	0.67	1	Anterior MI
0.05	0.15	0.00	70	0.8	1	Anterior MI
0.05	0.2	0.00	71	0.83	1	Anterior MI
0.1	0.29	0.00	76	0.6	1	Anterior MI
0.03	0.15	0.00	80	0.61	1	Anterior MI
0.1	0.2	0.00	80	0.61	1	Anterior MI
0.08	0.2	0.00	70	0.8	1	Anterior MI
0.05	0.2	0.00	72	0.67	1	Anterior MI
0.08	0.2	0.00	78	0.6	1	Anterior MI
0.03	0.25	0.00	76	0.6	1	Anterior MI
0.1	0.29	0.00	78	0.6	1	Anterior MI

Compilation of References

Adlam, D., & Hampton, R. (1997). *The ECG made easy*. Amsterdam: Elsevier.

Alexopoulos, E., Dounias, G. D., & Vemmos, K. (1999). Medical Diagnosis of Stroke Using Inductive Machine Learning. Machine Learning and Applications: Machine Learning in Medical Applications, 20-23.

Al-kazemi, B., & Mohan, C. K. (2002). Training feedforward neural network using multi-phase particle swarm optimization. *Proc 9th Intl Conf Neural Infor Process*, (2), 2615-2619. doi:10.1109/ICONIP.2002.1201969

Amato, F., Lopez, A., Maria, E., Vanhara, P., Hampl, A., & Havel, J. (2013). Artificial Neural Network in Medical Diagnosis. *Journal of Applied Biomedicine*, *11*(2), 47–58. doi:10.2478/v10136-012-0031-x

Arkin, R. (1998). *Behavior-Based Robotics*. Cambridge, MA: MIT Press.

Asadi, H., Dowling, R., Yan, B., & Mitchell, P. (2014). Machine learning for outcome prediction of acute ischemic stroke post intra-arterial therapy. *PLoS One*, *9*(2), e88225. doi:10.1371/journal.pone.0088225 PMID:24520356

Bai, Q. (2010). Analysis of particle swarm optimization algorithm. *Computer and Information Science*, *3*(1), 180–183. doi:10.5539/cis.v3n1p180

Behbahani, S. (2007). Investigation of adaptive filtering for noise cancellation in ECG signals. *2nd Intl Multi-Symp Comp Comput Sci*, 144–149.

Bekir, K. M., Osman, T., & Musa, A. (2003). A fuzzy clustering neural network architecture for multifunction upper-limb prosthesis. *IEEE Transactions on Biomedical Engineering*, *50*(11), 1255–1261. doi:10.1109/TBME.2003.818469 PMID:14619995

Bello, R., & Verdegay, J. L. (2012). Rough sets in the soft computing environment. *Information Sciences*, *212*, 1–14. doi:10.1016/j.ins.2012.04.041

Beni, G., & Wang, J. (1989). Swarm intelligence in cellular robotics systems. *Proc. NATO Advanced Workshop on Robots and Biological System*, 102.

Bergh, F. V. D. (2001). *An analysis of particle swarm optimizers* (Thesis). University of Pretoria, Pretoria, South Africa.

Bergh, V. D. F. (1999). Particle swarm weight initialization in multi-layer perceptron artificial neural networks. Paper accepted for ICAI, Durban, South Africa.

Bergh, V. D. F., & Engelbrecht, A. P. (2000). Cooperative learning in neural networks using particle swarm optimizers. *South African Computer Journal*, *26*, 84–90.

Birkner, M. D., Kalantri, S., & Solao, V. (2007). Creating diagnostic scores using data-adaptive regression: An application to prediction of 30-day mortality among stroke victims in a rural hospital in India. *Therapeutics and Clinical Risk Management*, *3*, 475–484. PMID:18488068

Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford, UK: Oxford University Press.

Bishop, C. M. (2007). *Pattern recognition and machine Learning*. Information Science and Statistics.

Blondin, J. (2009). *Particle Swarm Optimization: A Tutorial*. Amsterdam University.

Bonabeau, D., & Theraulaz, G. (1999). *Swarm intelligence: From natural to artificial system*. New York: Oxford University Press.

Bonabeau, E., Dorigo, M., & Theraulaz, G. (1999). *Swarm Intelligence: From Natural to Artificial Systems*. Oxford University Press.

Bonfa, I., Maioli, C., Sarti, F., Milandri, G. L., & Monte, P. R. D. (1993). *HERMES: An Expert System for the Prognosis of Hepatic Diseases*. Technical Report UBLCS-93-19. Universiti of Bologna.

Brause, R. W. (2001). Medical Analysis and Diagnosis by Neural Networks. J.W. Goethe-University.

Buczak, A., & Uhrig, R. (1996). Hybrid fuzzy-genetic technique for multisensory fusion. *Information Sciences*, *93*(3-4), 265–281. doi:10.1016/0020-0255(96)00078-3

Caldwell, B. E. Jr. (2007). *Prevent and Reverse Heart Disease*. Penguin Group.

Caruana, R., Baluja, S., & Mitchell, T. (1996). Advances in Neural Information Processing Systems: Vol. 8. *Using the Future to "Sort Out" the Present: Rankrop and Multitask Learning for Medical Risk Evaluation*. Cambridge, MA: The MIT Press.

Castro, B., Kogan, D., & Geva, A. B. (2000). ECG feature extraction using optimal mother wavelet. *The 21st IEEE Convention of the Electrical and Electronic Engineers*, 346-350. doi:10.1109/EEEI.2000.924422

Ceylan, R., Yuksel, O., & Bekir, K. A. (2009). Novel approach for classification of ECG arrhythmias: Type-2 fuzzy clustering neural network. *Expert Systems with Applications*, *36*(3), 6721–6726. doi:10.1016/j.eswa.2008.08.028

Chan, L.-W. (1999). Weighted least square ensemble networks. *Proceedings of the IJCNN*, *99*, 1393–1396.

Chazal, P., & Celler, B. G. (1998). Selecting a neural network structure for ECG diagnosis. *20th Ann Intl Conf IEEE – EMBS*, 1422-1425.

Chellappa, M. (1995). Telemedic-Care. In *NCIT'95: 8'th National Conference Information Technology'95*. Gabungan Komputer Nasional Malaysia.

Chun-an, L. (2008). New dynamic constrained optimization PSO algorithm. In *Proc. Fourth International Conference on Natural Computation* (pp. 650-653). Academic Press.

Clancey, W. J., & Shortliffe, E. H. (1984). *Readings in medical artificial intelligence: The first decade*. Reading, MA: Addison Wesley.

Coiera, E. (2003). *Guide to Medical Informatics, the Internet and Telemedicine*. Academic Press.

Cong, L., Sha, Y., & Jiao, L. (2006). Numerical optimization using organisational particle swarm algorithm. Simulated Evolution and Training. *Lecture Notes in Computer Science*, (4247), 150–157. doi:10.1007/11903697_20

Corne, D., Dorigo, M., & Glover, F. (Eds.). (1999). New Ideas in Optimization. McGraw Hill.

Cui, S., & Weile, D. S. (2005, November). Application of a parallel particle swarm optimization scheme to the design of electromagnetic absorbers. *IEEE Transactions on Antennas and Propagation*, *53*(11), 3616–3624. doi:10.1109/TAP.2005.858866

Detmer, W. M., & Shortliffe, E. H. (1997). Using the Internet to Improve Knowledge Diffusion in Medicine. *Communications of the Associations of Computing Machinery*, *40*(8), 101–108. doi:10.1145/257874.257897

Devi, S., Jadhav, D. G., & Pattnaik, S. S. (2011). *PSO based memetic algorithm for unimodal and multimodal function optimization. In Swarm, Evolutionary, and Memetic Computing, LNCS, 7076* (pp. 127–134). Berlin: Springer.

Dong, H. K., Abraham, A., & Jae, H. C. (2007). A hybrid genetic algorithm and bacterial foraging approach for global optimization. *Information Sciences*, (177), 3918–3937.

Dorigo, M., Maniezzo, V., & Colorni, A. (1996). The ant system: Optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man, and Cybernetics*, *26*(1), 29–41. doi:10.1109/3477.484436 PMID:18263004

Eberhart, R. C., & Hu, X. (1999). *Human Tremor Analysis Using Particle Swarm Optimization*. Retrieved from www.swarmintelligence.org/papers/CEC1999Human.pdf

Eberhart, R., & Shi, Y. (2001). Particle swarm optimization: developments, application and resources. *Proceedings of IEEE Congress of Evolutionary Computation*, 81-86. doi:10.1109/CEC.2001.934374

Eberhart, R., Shi, Y., & Kennedy, J. (2001). *Swarm Intelligence*. Academic Press.

Ebrahim, Z. A., Khazaee, A., and Ranaee, V. (2010). Classification of the electrocardiogram signals using supervised classifiers and efficient features. *Comp Meth Prog Biomed*, (99), 179-194.

Elif, D. U. (2010). Recurrent neural networks employing Lyapunov exponents for analysis of ECG signals. *Expert Systems with Applications*, (37): 1192–1199.

Farmer, J. D., Packard, N. H., & Perelson, A. S. (1986). The immune system, adaptation and machine learning. *Physica D. Nonlinear Phenomena*, *22*(1-3), 187–204. doi:10.1016/0167-2789(86)90240-X

Ferguson, D. (2004). *Particle swarm*. University of Victoria.

Fitzgibbon, E., Berger, R., Tsitlik, J., & Halperin, H. R. (2002). Determination of the noise source in the electrocardiogram during cardiopulmonary resuscitation. *Critical Care Medicine*, *30*(4Supplement), 148–153. doi:10.1097/00003246-200204001-00006 PMID:11953642

Flake, G. (1999). *The computational beauty of nature*. Cambridge, MA: MIT Press.

Fnaiech, F., Abid, S., Ellala, N., & Cheriet, M. (2002). A comparative study of fast neural network learning algorithms. In *Proceedings of IEEE International Conference on Systems, Man and Cybernetics (IEEE SMC)* (pp. 24-29). IEEE. doi:10.1109/ICSMC.2002.1173379

Formato, R. A. (2008). Central force optimization: A new nature inspired computational framework for multidimensional search and optimization. *Studies in Computational Intelligence*, *129*, 221–238. doi:10.1007/978-3-540-78987-1_21

Fredric, M. H., & Soowhan, H. (1996). Classification of cardial arrhythmias. *IEEE Transactions on Biomedical Engineering*, *43*(4), 425–429. doi:10.1109/10.486263 PMID:8626192

Gacek, A., & Pedrycz, W. (2012). *ECG Signal Processing, Classification and Interpretation*. New York: Springer. doi:10.1007/978-0-85729-868-3

Hansen, L. K., & Salamon, P. (1990). Neural network ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *12*(10), 993–1001. doi:10.1109/34.58871

Hardy, A. (1996). On the number of clusters. *Computational Statistics & Data Analysis*, *23*(1), 83–96. doi:10.1016/S0167-9473(96)00022-9

Hassoun, M. H. (1995). *Fundamentals of artificial neural network*. Cambridge, MA: MIT.

Ho, K. C., Speier, W., & El-Saden, S. (2014). Predicting discharge mortality after acute ischemic stroke using balanced data. *AMIA Annu Symp Proc*, 1787–96.

Hoong, N. K. (1988). Medical Information Science - Framework and Potential. In *International Seminar and Exhibition Computerization for Development-the Research Challenge*. Universiti Pertanian Malaysia.

Hope, T. M., Seghier, M. L., Leff, A. P., & Price, C. J. (2013). Predicting outcome and recovery after stroke with lesions extracted from MRI images. *NeuroImage. Clinical*, *2*, 424–433. doi:10.1016/j.nicl.2013.03.005 PMID:24179796

Hopwood, S. J., & Jeans, J. (2009). *An Introduction to the Kinetic Theory of Gases*. Cambridge University Press.

Hornik, K. M., Stinchcombe, M., & White, H. (1989). Multilayer feedforward, networks are universal approximators. *Neural Networks*, *2*(5), 359–366. doi:10.1016/0893-6080(89)90020-8

Hosseini, H. G., Reynolds, K. J., & Powers, D. (2001). A multi-stage neural network classifier for ECG events. *Proceedings 23rd Intl Conf IEEE EMBS*, (2), 1672-1675.

Hu, G. M., Palreddy, S., & Tompkins, W. (1997). Patient adaptable ECG beat classifier using a mixture of experts approach. *IEEE Transactions on Biomedical Engineering*, (44): 891–900. PMID:9282481

Idoumghar, L., Melkemi, M., Schott, R., & Aouad, M. I. (2011). *Hybrid PSO-SA type algorithms for multimodal function optimization and reducing energy consumption in embedded systems. Applied Computational Intelligence and Soft Computing*. Hindawi.

Inan, O. T., Giovangrandi, L., & Kovacs, G. T. A. (2006). Robust neural network based classification of premature ventricular contractions using wavelet transform and timing interval features. *IEEE Transactions on Biomedical Engineering*, *53*(12), 2507–2515. doi:10.1109/TBME.2006.880879 PMID:17153208

Ismail, W., & Shamsuddin, S. M. (2008). *Differential evolution for neural networks training enhancement* (M.Sc. Thesis). University Technology of Malaysia.

Jankowski, N. (1999). Approximation and Classification in Medicine with IncNet Neural Networks. In Machine Learning and Applications: Machine Learning in Medical Applications, 53-58.

Jesmin, K., Sharif, B., & Gregory, M. (2011). Empirical mode decomposition of the ECG signal for noise removal. *Optical Pattern Recognition*, (8055), 504-512.

Jones, M. T. (2005). *AI Application Programming* (2nd ed.). Hingham, MA: Charles River Media Inc.

Jorand, M. I., & Bishop, C. M. (1996). *Neural networks* (Technical Report No. A. I. Memo No. 1562). Artificial Intelligence Laboratory.

Karabatak, M., & Cevdet-Ince, M. (2009). An expert system for detection of breast cancer based on association rules and neural network. *Expert Systems with Applications*, *36*(2), 3465–3469. doi:10.1016/j.eswa.2008.02.064

Karakülah, G., Dicle, O., & Koşaner, O. (2014). Computer based extraction of phenoptypic features of human congenital anomalies from the digital literature with natural language processing techniques. *Studies in Health Technology and Informatics*, *205*, 570–574. PMID:25160250

Karl, G. R., Isis, A. W., Roberto, L., & Hakan, N. (2004). Fetal ECG waveform analysis. Best Practice & Research. *Clinics in Obstetrics and Gynaecology*, *3*(18), 485–514.

Kennedy, J., & Eberhart, R. (1995). A new optimiser using particle swarm theory. *Proceedings of the 6th International Symposium on Micro Machine and Human Science*, 39-43.

Khan, J., Wei, J. S., Ringnér, M., Saal, L. H., Ladanyi, M., Westermann, F., ... Meltzer, P. S. (2001). Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nature Medicine*, *7*(6), 673–679. doi:10.1038/89044 PMID:11385503

Kim, G. H., Yoon, J. E., An, S. H., Cho, H. H., & Kang, K. I. (2004). Neural network model incorporating: A genetic algorithm in estimating construction cost. *Building and Environment*, *39*(11), 1333–1340. doi:10.1016/j.buildenv.2004.03.009

Kirkpatrik, S., Gellat, C. D., & Vecchi, M. P. (1983). *Optimization by simulated annealing*. Academic Press.

Kohavi, R., & Provost, F. (1998). Glossary of Terms. *Machine Learning*, *30*(2/3), 271–274. doi:10.1023/A:1017181826899

Kononenco, I. (2001). Machine Learning for Medical Diagnosis: History, State of the Art and perspective. *Artificial Intelligence in Medicine*, *23*(1), 89–109. doi:10.1016/S0933-3657(01)00077-X PMID:11470218

Konstantinos, N. P., Dimitrios, H., & Jimmy, K. M. L. (2006). *ECG biometric recognition without fiducial detection, Biometric Consortium Conference*. IEEE Biometrics Symp, Baltimore, MD.

Konstantinos, N. P., Dimitrios, H., & Jimmy, K. M. L. (2006). *ECG biometric recognition without fiducial detection*. Biometric Consortium Conference. IEEE Biometrics Symposium, Baltimore, MD.

Krogh, A., & Vedelsby, J. (1995). Neural network ensembles, cross validation, and active learning. *Advances in Neural Information Processing Systems*, *7*, 231–238.

Krose, B., & Smagt, V. D. P. (1996). *An introduction to neural networks*. The University of Amsterdam.

Kumar, A. V. S. (2012). Diagnosis of Heart Diseases Using Fuzzy Resolution Mechanism. *Journal of Artificial Intelligence*, *5*(1), 47–55. doi:10.3923/jai.2012.47.55

Leonard, B. L. (2004). *The Kinetic Theory of Gases*. Dover Phoenix Editions.

Lian, Y., & Hoo, P. C. (2006). Digital elliptic filter application for noise reduction in ECG signal. *WSEAS Trans Electron*, *3*(1), 65–70.

Lian, Z., Zhu, F., Guan, Z., & Shao, X. (2008). The analysis of particle swarm optimization algorithm's convergence. *Proceedings of the 7th World Congress on Intelligent Control and Automation*, 623-628.

Ling, B. W. K., Ho, C. Y. F., Lam, H. K., Wong, T. P. L., Chan, A. Y. P., & Tam, P. K. S. (2008). Fuzzy rule based multiwavelet ECG signal denoising. *IEEE World Cong Comput Intell*, 1064-1068.

Lin, H., Wensheng, H., Xiaolin, Z., & Chenglin, P. (2006). Recognition of ECG patterns using artificial neural network. *Proceedings of the Sixth IEEE Intl Conf Intell, Sys Design and Appl*, 477-481.

Liu, H., & Motoda, H. (1998). *Feature Extraction, Construction and Selection: A Data Mining Perspective*. Kluwer Academic. doi:10.1007/978-1-4615-5725-8

Losada, R. A. (2004). Design finite impulse response digital filters. Part II. *Microwaves & RF*, (43): 70–84.

Lotfi, A. Z. (1969). Biological application of the theory of fuzzy sets and system. Biocybernetics of the Central Nervous System, 199-212.

Love, A., Arnold, C. W., & El-Saden, S. (2013). Unifying acute stroke treatment guidelines for a bayesian belief network. *Studies in Health Technology and Informatics*, *192*, 1012. PMID:23920786

Machado, L. O. (1996). *Medical Applications of Artificial Neural Networks: Connectionist Model of Survival* (Ph.D dissertation). Stanford University.

Mahabala, H. N., Chandrasekhara, M. K., Baskar, S., Ramesh, S., & Somasundaram, M. S. (1992). ICHT: An Intelligent Referral System for Primary Child Health Care. *Proceedings SEARCC*, *92*(11).

Mahmoodabadi, S. Z., Ahmadian, A., & Abolhasani, M. D. (2005). ECG feature extraction using daubechies wavelets. *Proceedings of the fifth IASTED International conference on Visualization, Imaging and Image Processing*, 343-348.

Mahmoudi, T. M., Forouzideh, N., Lucas, C., & Taghiyareh, F. (2009). Artificial neural network weights optimization based on imperialist competitive Algorithm. *Proceedings of the Seventh International Conference on Computer Science and Information Technologies*, 118-121.

Manickam, S., & Abidi, S. S. R. (1999). Experienced Based Medical Diagnostics System Over The World Wide Web (WWW). *Proceedings of The First National Conference on Artificial Intelligence Application In Industry*, 47 - 56.

Manolescu, D. A. (2000). *Feature Extraction—A Pattern for Information Retrieval*. Retrieved from https://micro-workflow.com/PDF/plop98.pdf

Manolescu, D. A. (n.d.). *Feature Extraction—A Pattern for Information Retrieval*. Academic Press.

Mehmet, E. (2004). ECG beat classification using neuro-fuzzy network. *Patt Recog Lett*, *25*(15), 1715–1722. doi:10.1016/j.patrec.2004.06.014

Meng, Y. K. (1996). Interval-Based Reasoning in Medical Diagnosis. *Proceedings of National Conference on Research and Development in Computer Science and Its Applications (REDECS'96)*, 220 - 224.

Minami, K., Nakajima, H., & Toyoshima, T. (1999). Real-time discrimination of ventricular tachyarrhythmia with Fourier transform neural network. *IEEE Transactions on Biomedical Engineering*, *46*(2), 179–185. doi:10.1109/10.740880 PMID:9932339

Moein, S., Monadjemi, S. A., & Moallem, P. A. (2008). Novel fuzzy-neural based medical diagnosis system. *Int J. Biological and Life Sciences*, *4*(3), 146–150.

Murali, N. V. (2004). *Design of FIR filters with arbitrary responses. Seminar report*. Bombay: Electrical System Group.

Murdoch, T. B., & Detsky, A. S. (2013). The inevitable application of big data to health care. *Journal of the American Medical Association*, *309*(13), 1351–1352. doi:10.1001/jama.2013.393 PMID:23549579

Murff, H. J., FitzHenry, F., & Matheny, M. E. (2011). Automated identification of postoperative complications within an electronic medical record using natural language processing. *Journal of the American Medical Association*, *306*, 848–855. PMID:21862746

Naghsh-Nilchi, A. R., & Kadkhodamohammadi, A. R. (2008). *Cardiac arrhythmias classification method based on MUSIC morphological descriptors, and neural network. EURASIP J Adv Sig Process*.

Neill, D. B. (2013). Using artificial intelligence to improve hospital inpatient care. *IEEE Intelligent Systems*, *28*(2), 92–95. doi:10.1109/MIS.2013.51

Omran, M. G. H. (2004). *Particle Swarm Optimization Methods for Pattern Recognition and Image Processing*. Retrieved from https://pdfs.semanticscholar.org/f55a/20f7f7a6ba00ecbd958be6cfb2c60324fe06.pdf

Orfanidis, S. J. (1996). *Introduction to Signal Processing*. Upper Saddle River, NJ: Prentice Hall.

Osowski, S., & Linh, T. H. (2001). ECG beat recognition using fuzzy hybrid neural network. *IEEE Transactions on Biomedical Engineering*, *48*(11), 1265–1271. doi:10.1109/10.959322 PMID:11686625

Owis, M. I., Youseff, A. B. M., & Kadah, Y. M. (2002). Characterisation of electrocardiogram signals based on blind source separation. *Medical & Biological Engineering & Computing*, *40*(5), 557–564. doi:10.1007/BF02345455 PMID:12452417

Ozbay, Y., Ceylan, R., & Karlik, B. (2006). A fuzzy clustering neural network architecture for classification of ECG arrhythmias. *Computers in Biology and Medicine*, *36*(4), 376–388. doi:10.1016/j.compbiomed.2005.01.006 PMID:15878480

Pareja-Tobes, P., Pelta, D., Sancho-Royo, A., & Verdegay, J. L. (2008). Search spaces representation in optimization problems. *Expert Systems with Applications*, *34*(4), 2891–2895. doi:10.1016/j.eswa.2007.05.046

Patel, V. L., Shortliffe, E. H., Stefanelli, M., Szolovits, P., Berthold, M. R., Bellazzi, R., & Abu-Hanna, A. (2009). The coming of age of artificial intelligence in medicine. *Artificial Intelligence in Medicine*, *46*(1), 5–17. doi:10.1016/j.artmed.2008.07.017 PMID:18790621

Paul, J. S., Reddy, M. R. S., & Kumar, V. J. (1997). Automatic detection of PVC's using autoregressive models. *19th Ann Intl Conf IEEE EMBS*, 68-71.

Pelta, D., Krasnogor, N., Bousono-Calzon, C., Verdegay, J. L., Hirst, J., & Burke, E. K. (2005). A fuzzy sets based generalization of contact maps for the overlap of protein structures. *Journal of Fuzzy Sets and Systems*, (152), 103-123.

Peng, L., Yang, B., & Jiang, J. (2009). A novel feature selection approach for biomedical data classification. *Journal of Biomedical Informatics*, *179*(1), 809–819. doi:10.1016/j.jbi.2009.07.008 PMID:19647098

Physionet. (2012). *Physiologic signal archives for biomedical research, A database for heart signals*. Retrieved February 20, 2010 from www.physionet.org/physiobank

Physionet. (n.d.). *Physiologic signal archives for biomedical research, A database for heart signals*. Retrieved from www.physionet.org/physiobank

Pomi, A., & Olivera, F. (2006). *BMC Medical Informatics and Decision Making, Context-sensitive auto associative memories as expert systems in medical diagnosis*. BioMed Central.

Poungponsri, S., & Yu, X. H. (2009). Electrocardiogram (ECG) signal modelling and noise reduction using wavelet neural networks. *Proceedings of the IEEE Intl Conf Autom Logistics*, 394-398.

Raj, V. N. P., & Venkateswarlu, T. (2011). ECG signal denoising using undecimated wavelet transform. *Electronics Computer Technology, 3rd International Conference on*, 94-98 doi:10.1109/ICECTECH.2011.5941808

Rashedi, E., Nezamabadi-pour, H., & Saryazdi, S. (2009). GSA: A gravitational search algorithm. *Information Sciences*, *179*(13), 2232–2248. doi:10.1016/j.ins.2009.03.004

Robert, M., & Ian, R. G. (2006). The effects of noise on computerised electrocardiogram measurements. *Journal of Electrocardiology*, *39*(4), 165–173. doi:10.1016/j.jelectrocard.2006.05.025

Roger, A., & Hans-Jakob, S. (2009). Meet the challenge of high-pass filter and ST-segment requirements with a DC-coupled digital electrocardiogram amplifier. *Journal of Electrocardiology*, *46*(6), 574–579. PMID:19700169

Salerno, J. (1997). Using the particle swarm optimization technique to train a recurrent neural network. *Proceedings of ninth IEEE international conference on tools with artificial intelligence*, 45-49. doi:10.1109/TAI.1997.632235

Sancho-Royo, A., Pelta, D., Verdegay, J. L., & Gonzalez, J. R. (2005). Evaluation of co-operative strategies in optimization problems. *Proc. Nisis*. Retrieved from http://www.nisis.de

Sarle, W. S. (1999). *Neural network FAQ, part 1 of 7: Introduction*. Retrieved from ftp://ftp.sas. com/pub/neurl/FAQ.html

Senhadji, L., Carrault, G., Bellanger, J. J., & Passariello, G. (1995). Comparing wavelet transforms for recognising cardiac patterns. *IEEE Eng Med Biol*, *14*(2), 167–173. doi:10.1109/51.376755

Sharkey, S. W., Lesser, J. R., Zenovich, A. G., Maron, M. S., Lindberg, J., & Longe, T., F., & Maron, B. J. (2005). Acute and reversible cardiomyopathy provoked by stress in women from the United States. *Circulation, 111*, 472–479. doi:.CIR.0000153801.51470.EB10.1161/01

Shaw, E. (1962). The schooling of fishes. *Scientific American*, *206*(6), 128–138. doi:10.1038/scientificamerican0662-128 PMID:14458553

Shortliffe, E. H. (1998). *The Evalution of Health-Care Records in the Era of the Internet*. Retrieved from http://smi-web.stanford.edu/pubs/abstracts_by_author/

Shortliffe, E. H. (1987). Computer Programs to Support Clinical Decision Making. *Journal of the American Medical Association*, *258*(1), 61. doi:10.1001/jama.1987.03400010065029 PMID:3586293

Shortliffe, E. H., & Buchanan, B. G. (1975). A Model of Inexact Reasoning in Medicine. *Mathematical Biosciences*, *23*(3-4), 351–379. doi:10.1016/0025-5564(75)90047-4

Shyu, L. Y., Wu, Y. H., & Hu, W. C. (2004). Using wavelet transform and fuzzy neural network for VPC detection from the Holter ECG. *IEEE Transactions on Biomedical Engineering*, *51*(7), 1269–1273. doi:10.1109/TBME.2004.824131 PMID:15248543

Silver, M. (2002). *Success With Heart Failure: Help and Hope for Those with Congestive Heart Failure* (3rd ed.). Da Capa Press.

Sivanandam, S. N., & Deepa, S. N. (2007). Introduction to genetic algorithms. Springer.

Sonoda, S., & Murata, N. (2013). *Nonparametric weight initialization of neural networks via integral representation*. Retrieved from http://arxiv.org/abs/1312.6461

Sordo, M. (2002). *Introduction to Neural Networks in Healthcare. OpenClinical: Knowledge Management for Medical Care*. Harvard.

Sornmo, L., & Laguna, P. (2005). *Bioelectrical Signal Processing in Cardiac and Neurological Applications*. Academic Press.

Sotos, J. M., Melendez, C. S., Salort, C. V., Abad, R. C., and Ibanez, J. J. R. (2007). A learning based Widrow-Hoff Delta algorithm for noise reduction in biomedical signals, Bio-inspired Modelling of Cognitive Tasks. *Lecturer Notes in Computer Science*, (4527), 377-386.

Sotos, J. M., Sanchez, C., Mateo, J., Alcaraz, R., Vaya, C., & Rieta, J. J. (2007). Neural networks based approach to remove baseline drift in biomedical signals. *11th Mediterr Conf Med Biomed Eng Comput 2007. IFMBE Proceedings*, *16*(2), 90–93.

Street, W. N., Mangasarian, O. L., & Wolberg, W. H. (1996). Individual and Collective Prognostic Prediction. *Thirteenth International Conference on Machine Learning*.

Tadejko, P., & Rakowski, W. (2007). Mathematical morphology based ECG feature extraction for the purpose of heartbeat classification. *6th International Conference on Computer Information Systems and Industrial Management Applications*, 322-327. doi:10.1109/CISIM.2007.47

Taniguchi, M., & Tresp, V. (1997). Averaging regularized estimator. *Neural Computation*. doi:10.1162/neco.1997.9.5.1163

Thakur, A., Mishra, V., & Jain, S. (2011). Feed forward artificial neural network: Tool for early detection of ovarian cancer. *Scientia Pharmaceutica*, *79*(3), 493–505. doi:10.3797/scipharm.1105-11 PMID:21886899

Theodorou, T., & Ketikidis, P. (1995). Neo-Dat An Expert System to Support the Designers of Clinical Trails. *5th Hellenic Conference on Informatics*.

Tomas, B. G. (n.d.). Retrieved from http://www.amazon.co.uk/Introduction-12-Lead-ECG-Interpretation-Garcia/dp/0763719617

Tomas, B. G., & Neil, E. H. (2002). *Introduction to 12-Lead ECG: The Art of Interpretation*. Jones and Bartlett Publishers.

Übeyli, E. D. (2007). Implementing automated diagnostic systems for breast cancer detection. *Expert Systems with Applications*, *33*(4), 1054–1062. doi:10.1016/j.eswa.2006.08.005

Van, D. B., & Engelbrecht, A. P. (2000). Cooperative learning in neural networks using particle swarm optimisers. *South African Comput*, (26), 84-90.

Venter, G., & Haftka, R. T. (2010). Constrained particle swarm optimization using a bi-objective formulation. *Structural and Multidisciplinary Optimization*, (40), 65-76. Retrieved from http://www.springerlink.com/content/1615-147x/40/1-6/

Wan, H. I. W., & Siraj, F. (2006). *Artificial intelligence in medical applications: an exploration*. Universiti Utara Malaysia.

Warner, D. (1997). *Malaysian Medical Matrix: Telemedicine in the age of the Multimedia Super Coridor*. Academic Press.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, *1*(6), 80–83. doi:10.2307/3001968

Willems, J. L., Arnaud, P., Van Bemmel, J. H., Bourdillon, P. J., Degani, R., Denis, B., ... Zywietz, C. (1987). A reference data base for multi-lead electrocardiographic computer measurement programs. *Journal of the American College of Cardiology*, *10*(6), 1313–1321. doi:10.1016/S0735-1097(87)80136-5 PMID:3680801

Yana, H., Jiangb, Y. T., Zhenge, J., Pengc, C., & Lid, Q. (2006). A Multilayer Perceptron-Based Medical Decision Support System for Heart Disease Diagnosis. *Expert Systems with Applications*, *2*(30), 272–281. doi:10.1016/j.eswa.2005.07.022

Yao, X., & Liu, Y. (1997). A new evolutionary system for evolving artificial neural networks. *IEEE Transactions on Neural Networks*, *8*, 694–713. doi:10.1109/72.572107

Yao, X., Liu, Y., & Lin, G. (1999). Evolutionary programming made faster. *IEEE Transactions on Evolutionary Computation*, (3): 82–102.

Yeap, T. H., Johnson, F., & Rachniowski, M. (1990). ECG beat classification by a neural network. *Proc Ann Intl Conf IEEE EMBS*,1457-1458. doi:10.1109/IEMBS.1990.691837

Ye, H., Shen, H., & Dong, Y. (2017). *Using Evidence-Based medicine through Advanced Data Analytics to work toward a National Standard for Hospital-based acute ischemic Stroke treatment*. Mainland China.

Ying, P. C., & Pei, J. (2010). Analysis of particle interaction in particle swarm optimization. *Theoretical Computer Science*, (411): 2101–2115.

Yu, J., Wang, S., & Xi, L. (2008). Evolving artificial neural networks using an improved PSO and DPSO. *Neurocomputing*, *71*(4-6), 1054–1060. doi:10.1016/j.neucom.2007.10.013

Zelic, I., Lavrac, N., Najdenov, P., & Rener-Primec, Z. (1999). Impact of machine Learning of the Diagnosis and Prognosis of First Cerebral Paroxysm. Machine Learning and Applications: Machine Learning in Medical Applications, 24-26.

Zhang, Q., Xie, Y., & Ye, P. (2013). Acute ischaemic stroke prediction from physiological time series patterns. *The Australasian Medical Journal*, *6*(5), 280–286. doi:10.4066/AMJ.2013.1650 PMID:23745149

Zhan, Z. H., Zhang, J., Li, Y., & Chung, H. S. H. (2009). Adaptive Particle Swarm Optimization. *IEEE Transactions on Systems, Man, and Cybernetics. Part B, Cybernetics*, *39*(6), 1362–1381. doi:10.1109/TSMCB.2009.2015956 PMID:19362911

Zhao, Q., & Zhan, L. (2005). ECG feature extraction and classification using wavelet transform and support vector machines. *International Conference on Neural Networks and Brain*, (2), 1089-1092.

Ziarani, A. K., & Konrad, A. (2004). A nonlinear adaptive method of elimination of powerline interferences in ECG signals. *IEEE Transactions on Biomedical Engineering*, *49*(6), 540–547. doi:10.1109/TBME.2002.1001968 PMID:12046699

Zne, J. L., & Chou, Y. L. (2005). A hybrid search algorithm with heuristics for resource allocation problem. *Informatics and Computer Science*, (173), 155-167.

Zrimec, T., & Kononenko, I. (2004). *Feasibility Analysis of Machine Learning in Medical Diagnosis from Aura Images*. University of Ljubljana.

Zumray, D., & Tamer, O. (2001). ECG beat classification by a novel hybrid neural network. *Comp Meth Prog Biomed*, (66), 167-181.

About the Author

Sara Moein, PhD, is currently a researcher in the department of Genetics and Genomic Sciences at Mount Sinai School of Medicine, United States. Her interests are machine learning, algorithm designing, optimization and computational biology. She has received her PhD from Multimedia University, Malaysia in computer science. Her Master and Bachelor degrees are in software engineering. She is author of book *Medical Diagnosis Using Artificial Neural Networks*. Dr. Sara Moein is the member of editorial boards of some of the international journals such as *Journal of Experimental & Theoretical Artificial Intelligence* and *Journal of Intelligent Automation & Soft Computing* and others. In addition, she is reviewer of many journals and conferences papers such as Journal of Computing, Journal of Supercomputing and conferences such as ICINCO (2012-2016), WORLDCOMP (2009-2013) and 7th IEEE BIBE. She has a number of publications in book chapters, journals, and conference proceedings.

Index

I

J

K

L

M

N

O

P

R

S

T

U

V

W