AN INTEGRATED FUZZY MODEL FOR PATTERN RECOGNITION

by

ABDU MASANAWA SAGIR

Thesis submitted in the fulfilment of the requirements
for the degree of
Doctor of Philosophy

February 2017

ACKNOWLEDGEMENT

My profound gratitude goes to Almighty Allah, the All-Knowing, who gave me this opportunity to pursue my PhD Degree program.

My special gratitude goes to my supervisor Associate Professor Dr. Saratha Sathasivam for her patience, motivation, constant encouragement, who deeply contributes valuable guidance and suggestions to the success of my PhD Degree research program.

I am highly indebted to express my sincere gratitude to Dean, Professor Dr. Hailiza Binti Kamarulhaili, academic and non-academic staff of School of Mathematical Sciences, Universiti Sains Malaysia.

It is a great pleasure for me to acknowledge Hassan Usman Katsina Polytechnic in collaboration with Tertiary Education Trust Funds (TETFund) of Nigeria for sponsoring my PhD studies under the Academic Staff Training and Development Scheme.

I would like to thanks my colleagues whom I cannot explain how I enjoyed staying with them at Universiti Sains Malaysia.

I greatly appreciate to my siblings Alh. Ibrahim Sagir, Alh. Nasiru Sagir, Alh. Shehu Sagir, Haj. Zulai Sagir, Haj. Turai Sagir, Haj. Amina Sagir and Haj. Nana Sagir, to mention few, for their kind supports, prayers and advices.

I would not forget to extend my sincere and innermost gratitude to my friends and well-wishers at home who contributed to me in one way or the other, especially "Madahan Katsina" Alh. Mustapha Madaha Gafai, Alh. Sule Falalu Rafindadi, Alh. Tanimu Ibrahim Darma, Alh. Yusuf Amadu, "Sarkin Dawakin Tsakar Gidan Nassarawa – Katagum Emirate Council" Prince Alh. Halilu Kabir Umar, Lawal Maikudi, Dr. Mukhtar Abubakar Balarabe, Muhammad Zayyad Umar, Alh. Kabir Falalu, Yusuf Mohammed D, and Alh. Kabir Dodo.

I wish to dedicate this PhD thesis to my deepest and most heart-felt late parents Alh. Sagir Masanawa and Nana Bilkisu, my wife Haj. Amina and my children Abubakar, Nana Firdausi, Umar and Hajara.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABC	Artificial Bee Colony
ACO	Ant Colony Optimisation
ANFIS	Adaptive Neuro Fuzzy Inference System
ANN	Artificial Neural Network
CAD	Coronary Artery Dataset
CBR	Case Based Reasoning
CBRPSO	Case Based Reasoning Particle Swarm Optimisation
CE	Classifier Ensemble
DR	Diabetic Retinopathy
DTS	Dataset
ELM	Extreme Learning Machine
FAR	False Acceptance Rate
FN	False Negative
FP	False Positive
FRR	False Rejection Rate
FLANN	Functional Link Artificial Neural Network
FCM	Fuzzy C-Means
FES	Fuzzy Expert System
FIS	Fuzzy Inference System
FL	Fuzzy Logic
GFNN-GA	Generalized Feedforward Neural Network with Genetic Algorithm
GA-BPN	Genetic Algorithm with Back Propagation Network
GA-CFS	Genetic Algorithm with Classification by Feature Selection

GA-RS	Genetic Algorithm-Rough Sets
GA-AWAIS	Genetic Algorithms for Attribute Weighted Artificial Immune System
IFM	Integrated Fuzzy Model
KMP-PSO	Kernel matching pursuit with Particle Swarm Optimisation
KNN	Kohonen Neural Network
LSE	Least Square Error
LM	Levenberg-Marquardt algorithm
LMAD	Levenberg-Marquardt algorithm with Analytic Derivation
MATLAB	Matrix Laboratory
MF	Membership Function
MLM	Modified Levenberg-Marquardt algorithm
MLP	Multi Layer Perceptron
N_i	Number of inputs
Np	Number of parameters
Nr	Number of rules
Nt	Number of training data
PSO	Particle Swarm Optimisation
PCA-GA	Principal Component Analysis with Genetic Algorithm
RMSE	Root Mean square Error
SRLPSO	Self-Regulated Learning capability of Particle Swarm Optimisation
SANN	Simple Artificial Neural Network
SVD	Singular Value Decomposition
SVM	Support Vector Machine

TSK	Takagi Sugeno Kang
TN	True Negative
TP	True Positive
UCI	University of Califonia at Irvine
WEKA	Waikato Environment for Knowledge Analysis

LIST OF SYMBOLS

$\mathbf{a}_i$	ith row of matrix $\mathbf{A}$
$\mathbf{A}^\mathrm{T}$	Transpose of matrix $\mathbf{A}$
c, m	Centres of Gaussian membership functions
E	Error function
$\mathbf{g}$	Error gradient vector
$\mathbf{H}$	Hessian matrix
$\mathbf{I}$	Sparse identity matrix
$\mathbf{J}$	Jacobian matrix
k	Index of iterations
l	Layers of ANFIS-LMAD
n	Number of input variable
p	Initial number of partition
$\|S1\|$	Antecedent parameters
$\|S2\|$	Consequent parameters
$\hat{\mathbf{v}}$	Index vector of unique membership functions
$\mathbf{W}$	Weight vector
$\mu(x)$	Degree of membership function
σ, β	Width of Gaussian membership functions
Ψ	Learning parameter

SATU MODEL FUZZY GABUNGAN UNTUK PENGENALPASTIAN POLA

ABSTRAK

Diagnosis perubatan adalah satu proses mengkaji keadaan penyakit, atau komplikasi perubatan yang menghuraikan tentang tanda-tanda dan simptom pesakit. Diagnosis perubatan membantu mendapatkan ciri-ciri berlainan yang mewakili pelbagai variasi penyakit. Keputusan tentang kehadiran atau ketiadaan penyakit adalah satu tugas yang mencabar kerana banyak tanda dan simptom adalah tidak spesifik; dan banyak ujian perlu dijalankan. Untuk mengenalpasti diagnosis analisis simptom dengan tepat, doktor mungkin memerlukan satu sistem diagnosis yang efisien yang mampu meramal dan mengklasifikasi keadaan pesakit. Tesis ini menghuraikan satu metodologi untuk membangunkan satu model fuzzy yang digabungkan dengan menggunakan aplikasi sistem inferens fuzzy neuro adaptif (ANFIS) yang boleh digunakan oleh para doktor untuk mempercepatkan lagi proses diagnosis. Pendekatan pilihan ciri telah digunakan untuk mengenalpasti dan mengeluarkan atribut-atribut yang tidak penting, tidak relevan dan berlebihan daripada data yang tidak menyumbang kepada ketepatan model ramalan. Kaedah yang dicadangkan menggunakan teknik pengesahan *Hold-out*, yang membahagikan set data latihan dan ujian kepada dua pertiga dan satu pertiga. Kaedah yang disarankan menggunakan teknik pembahagian kekisi untuk menangani tujuh atribut dan fungsi keahlian Gaussian dari metod konvensyional yang sudah sedia dipasang iaitu Matlab, yang menggunakan sebilangan kecil atribut input iaitu biasanya kurang dari lima. Untuk menjamin kekuatannya, dua belas set data yang ditanda-aras yang

diperolehi dari University of California di repositori pembelajaran Irvine telah digunakan dalam kajian ini. Model yang disarankan meramal label klasifikasi, dan perbandingan telah dibuatdi antara latihan penurunan gradien Matlab kepada latihan Levenberg-Marquardt yang telah diubah-suai. Tambahan pula, satu percubaan telah dibuat untuk mengkhususkan keberkesanan prestasi yang mengukur ketepatan, sensitiviti dan pengkhususan. Keputusan klasifikasi yang telah diperolehi dari kaedah yang disarankan boleh dibandingkan dan dalam beberapa keadaan, ia lebih baik dari keputusan yang diperolehi melalui metod konvensyional dan lain-lain metod metaheuristik yang berkaitan. Kaedah yang dicadangkan (ANFIS-LMAD) memaksimakan data yang diklasifikasi dengan tepat dan meminimakan bilangan pola yang tidak diklasifikasi dengan tepat. Dengan kata lain, model ANFIS-LMAD boleh meramal dengan tepat label-label klasifikasi pesakit yang mempunyai penyakit dan mereka yang tidak berpenyakit. Kajian ini, telah mencapai tiga tekik utama yang baru:yang pertama, satu algoritma pembelajaran baru, algoritma Levenberg-Marquardt yang telah diubah-suai telah digunakan menggantikan algoritma penurunan gradien. Matriks Jacob dibangunkan menggunakan skema derivasi analitikal dan peraturan rantaian, yang menangani masalah kesukaran komputasi Matriks Jacob menggantikan skema perbezaan finit. Yang kedua, fungsi keahlian unik indeks dalam vektor mengikut barisan menggunakan teknik vektor untuk mendapatkan keputusan prestasi klasifikasi yang efisien dan ia memperkenalkan kadar penumpuan yang lebih cepat. Ketiga, satu kaedah efektif menukar matriks Jacob kepada matriks Jacob jarang telah digunakan, yang mana ia menyumbang kepada pengurangan penggunaan memori atau ruang penyimpanan, memperbaiki masa pelaksanaan dan mempercepatkan lagi pemprosesan data.

AN INTEGRATED FUZZY MODEL FOR PATTERN RECOGNITION

ABSTRACT

Medical diagnosis is a process of investigating which medical condition, disease or disorder describes signs and symptoms of a patient. Medical diagnosis helps to obtain different features representing the different variation of the disease. The decision about presence or absence of diseases of patients is a challenging task because many signs and symptoms are non-specific; and many tests might be required. To recognise an accurate diagnosis of symptom analysis, the physician may need efficient diagnosis system that can predict and classify patient condition. This thesis describes a methodology for developing an integrated fuzzy model by utilising the application of adaptive neuro fuzzy inference system (ANFIS) that can be used by physicians to accelerate diagnosis process. Feature selection approach was used to identify and remove unneeded, irrelevant and redundant attributes from the data that do not contribute to the accuracy of a predictive model. The proposed method used Hold-out validation technique, which divides the training and test data sets into two-thirds to one-third, respectively. The proposed method uses grid partition technique to cope with seven input attributes and Gaussian membership functions than conventional method built-in Matlab, which uses small number of input attributes usually less than five. For robustness, twelve benchmarked datasets obtained from University of California at Irvine's (UCI) machine learning repository were used in this research. The proposed model predicts the classification labels, and comparison was made between the Matlab's built-in gradient descent training to the Modified

Levenberg-Marquardt training. In addition, an attempt was done to specify the effectiveness of the performance measuring accuracy, sensitivity and specificity. The obtained classification results of the proposed method are comparable and in some cases superior to those obtained by using conventional method and other related existing metaheuristic methods. The proposed model (ANFIS-LMAD) maximises the correctly classified data and minimise the number of incorrectly classified patterns. In other words, the ANFIS-LMAD model was able to correctly predict classification labels of patients who have diseases and those who do not have diseases. This research, has achieved three major novelty techniques. Firstly, a novel approach learning algorithm, the Modified Levenberg-Marquardt algorithm was used instead of gradient descent algorithm. The Jacobian matrix was built up using analytical derivation scheme and chain rule, which overcome the complexity for computation of Jacobian matrix instead of finite difference scheme. Secondly, indexing unique membership functions in a row-wise vector using a vectorisation technique to obtain efficient classification performance results and make faster convergence rate was introduced. Thirdly, an effective way for transforming Jacobian matrix into sparse Jacobian matrix was employed, which also contributes towards reducing memory consumption or storage space, improved execution time and accelerate the processing of the data.

CHAPTER 1

GENERAL INTRODUCTION

1.1 Introduction

Currently, most of the real world classification systems are considered to be so difficult in the developmental process, as they may involve numerous trade-off problems like computational cost, accuracy, storage space and execution time. They may be linear or non-linear, predictable or unpredictable. Developing of an integrated fuzzy system is of importance in almost all fields, especially in medical disease diagnosis, transportation, signal processing and telecommunication, and engineering (Sadek, 2007).

This chapter introduces and describes the theory that explains why the research problem under study exists. It also contains background of the study, problem statement, objectives of the study, research methodology, an overview and structure of the thesis.

1.2 Background of Study

As reported by the World Health Organization (WHO) (2014), out of the 38 million deaths are due to non-communicable diseases (NCDs) in 2012, more than 40% were premature, affecting people under 70 years of age. The majority of premature NCD deaths are preventable. As a result, NCDs are one of the major health and development challenges of the 21st century.

Diagnosis of most of these diseases is very expensive because many signs and symptoms are nonspecific as such tests are required for predictions, such as tests in blood pressure, cholesterol, chest pain, blood sugar and so on. This research involved diagnosis of different medical diseases such as heart attack and hepatitis diseases.

Heart attack disease remains the main cause of death worldwide. WHO estimated 17.5 million people died from cardiovascular diseases in 2012, representing 31% of all deaths globally. An estimate of about 16 million died below the age of 70, 82% of which occurred in low and middle income countries. About 7.4 million of the deaths were due to coronary heart disease, whereas 6.7 million were due to stroke (World Health Organization, 2015).

Hepatitis is an injury to the liver with inflammation of the liver cell. There are five main different types of hepatitis viruses, referred to as type A, B, C, D, E and possibly G. Hepatitis A and E are of acute types whereas Hepatitis B, C and D lead to chronic disease (Davis, 2016). These five types are of the greatest concern because of the burden of illness and death they cause and the potential for outbreaks and epidemic spread. Infection from these viruses results in approximately 1.45 million deaths each year (World Health Organization, 2015). Eighty one per cent (81%) of world's infants are vaccinated and protected from hepatitis B infection. Two million occurrences of Hepatitis B and Hepatitis C infection occur yearly through unsafe injections. These viruses are also transmitted through contaminated water and food, as well as by contact with blood or bodily fluids, through unsafe injections or transfusions. Infection also occurs from a mother to a child, or through sexual contact (WHO, 2015).

Physicians make use of computerised technologies to assist in diagnosis and give suggestions as medical diagnosis is full of uncertainty. The best and efficient technique for dealing with uncertainty is by applying soft computing techniques, incorporating of fuzzy logic and neural network, which are complementary to each other rather than opposing each other for system identification or recognition. This approach has the capability to be familiarised with patterns and get used to themselves in order to handle a changing environment (Buragohain, 2009). The essential contribution of neural networks is a methodology for identification, learning and adaptation, while the essential contribution of fuzzy logic is a methodology for computing with words which can deal with ideas or knowledge and little details about data.

Neuro fuzzy systems are multilayer connectionist networks that realize the basic elements and functions of traditional fuzzy logic decision systems. Nonetheless, the performance of the system depends on the accuracy of the model. Therefore it is of greatest importance to build a model which correctly reflects the behaviour of the system under consideration.

1.3 Problem Statement

Some of the real world problems involve multiple inputs and outputs. Outputs are usually variables that one wants to investigate, while inputs are variables that cause an effect on the output if some changes have been made to it. The underlying relationship between the inputs and outputs (factors and responses) is often unknown. A lot of effort has been spent on developing methods that are able to model the relationship (Kacprzyk, 2010).

It is important to develop new methodologies because as our traditional methods face growing challenges, especially with increasing non-response with high computational cost. To evaluate the results obtained, the statistical and machine learning approach is of most importance in medical diagnosis; as such performance measures must be employed.

The proposed model is a new hybrid approach for training the adaptive network based on fuzzy inference system (ANFIS). The approach incorporates hybrid learning algorithm least square estimates with the Modified Levenberg-Marquardt (MLM) algorithm.

This study describes 3 major novelty techniques: 1. the Modified Levenberg-Marquardt (MLM) algorithm was used instead of Gradient Descent (GD) algorithm. The Jacobian matrix was built up using analytical derivation scheme and chain rule to overcome the complexity for computation of Jacobian matrix instead of finite difference scheme. 2. Indexing unique membership functions in a row-wise vector using a vectorisation technique to obtain good classification performance and make faster convergence speed was introduced. 3. An effective way for transforming Jacobian matrix into sparse Jacobian matrix was employed, which also contributes towards reducing memory consumption or storage space, improved execution time and speed up the processing of the data. The related existing algorithms, such as gradient descent, genetic algorithm, particle swarm optimisation and artificial bee colony lack these new techniques.

The proposed model involved using the grid partition method for dividing the data space into rectangular sub-spaces, the higher the input variables are, the less accuracy is to be obtained because it generates rules by enumerating all possible combinations of membership functions of all inputs. But, the proposed method was able to produce more accurate results despite its large number of rules. Because of its new techniques and properly selected relevant feature attributes that were involved in this research work.

The present research was used in developing an integrated fuzzy model (IFM) for real world classification problems like the diagnosis of medical diseases. Twelve benchmark medical data sets for applicability of the model were used. Performance comparison of the proposed model was used based on machine learning process with other related existing based Genetic Algorithm (GA), Particle Swarm Optimisation (PSO) and Artificial Bee Colony (ABC) algorithms.

There are limitations associated with this study that constrain the generalisation of the results. The main limitations are the selection of twelve benchmark medical data sets from the UCI machine learning repository.

Other approaches were used in developing and testing the proposed model in this study. Steps to build a model are feature selection, feature space, learning algorithm and performance error, whereas for testing of the model involves validation method, misclassification rate and performance measuring accuracy, sensitivity and specificity.

1.4 Objective of the Study

The main focus of this research is to develop an integrated fuzzy model for a real world problem system based on accessible record of input–output data using ANFIS and with the help of adapting a new and efficient hybrid learning algorithm. It was observed that in the conventional ANFIS models, most of the researchers relied on the use of a gradient descent algorithm for training of ANFIS.

The specific research objectives are:

- To develop an integrated fuzzy model that incorporates the capability of artificial neural network (ANN) and fuzzy logic (FL) for data classification purpose using a novel approach learning algorithm MLM for training ANFIS model.
- To improve the classification performance and convergence speed of the proposed model in terms of a vectorisation and sparse storage techniques.

- To demonstrate the applicability of the proposed model (ANFIS-LMAD) in data classification using twelve benchmark data sets in the areas of medical disease diagnosis.

1.5 Research Methodology

This section described the direction of research methodology adopted using flowchart to enable reader to cope with the underlying steps. Figure 1.1 shows the flowchart of research methodology.

Figure 1.1 provides a framework for the conduct of research methodology. It involves a simplified representation that explain the process of the proposed model development like procedures used for data extraction, extracting of initial model, data pre-processing, validation method, learning algorithm and performance evaluation.

The extracted data sets are obtained from University of California at Irvine's machine learning repository. For robustness twelve data sets is to be used for the diagnosis of medical diseases.

Extracting the initial model allow the method to select the most important input variables for the final model for the purpose of adjusting the set of network parameters in order to minimise errors between the network output and the given data.

Data pre-processing involves conversion of raw data into an understandable format. Data pre-processing involves (i) feature selection of the subset of the original or existing features without transformation, and (ii) missing data usually called missing values are nearly universal in statistical practice. Outlining missing values can help in defining the nature of the problem.

Validation method is the process of determining the extent to which a model is an accurate real world representation from the perspective of the intended uses of the model. In the proposed method, the stability of performance is to be measured by dividing the data set into training and testing with Hold out validation method.

Choosing proper neural network learning algorithm for adaptive based classifier problem is of most important as it serves a methodology for identification, learning and adaptation. The learning algorithm makes it suitable for problems whose structure is relatively unknown, such as pattern recognition, medical diagnosis, control system, time series prediction and so on. In the proposed method, Levenberg-Marquardt algorithm was modified and used for the purpose of matching the adaptive neuro fuzzy inference system based Levenberg-Marquardt algorithm with analytical derivation for computation of Jacobian matrix (ANFIS-LMAD) output with the training data.

Performance evaluation determines how well the models are worked after training. In this research, some approaches are employed in order to test the performance of the ANFIS-LMAD model, such as performance error, machine learning approach for computing the performance of metric and misclassification rate.

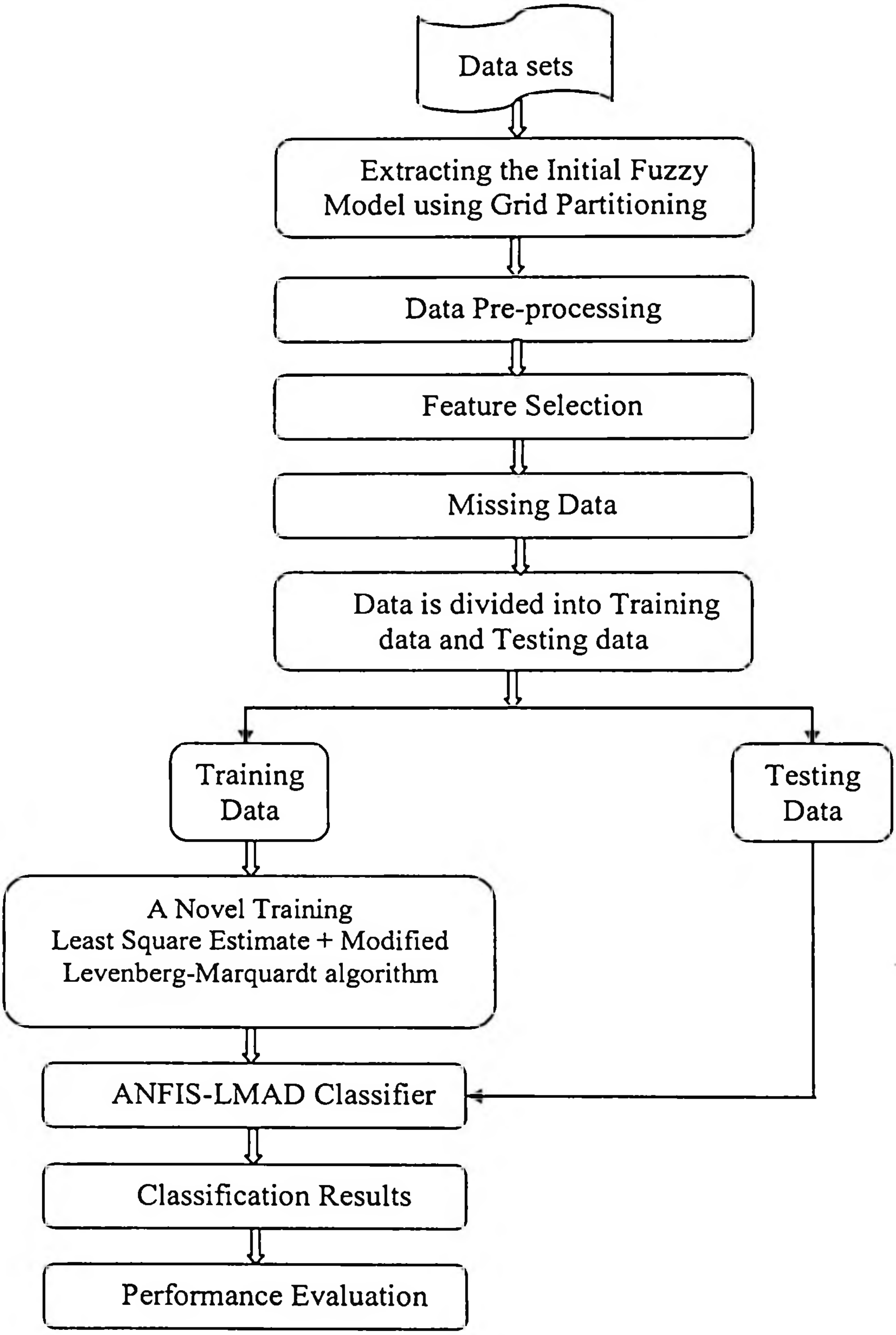

Figure 1.1 Flowchart of research methodology

1.6 Overview of the Thesis

Specific algorithms of this thesis are described in this section; such as Gradient-based descent algorithms (Steepest Descent (SD) and Levenberg-Marquardt algorithm) and Meta-heuristic algorithms (Genetic algorithm, Particle Swarm Optimisation and Artificial Bee Colony).

1.6.1 The Levenberg-Marquardt algorithm

The Levenberg-Marquardt (LM) algorithm, which was developed independently by Kenneth Levenberg and Donald Marquardt (Levenberg, 1944; Marquardt, 1963), suggest using a damped Gauss-Newton method (Madsen, Nielsen, & Tingleff, 2004).

The LM algorithm is one of the most widely used optimisation algorithms. It outperforms steepest descent and other evolutionary algorithms like the genetic algorithm, in terms of producing numerical solutions to the problem of minimising a non-linear function. A linear function is function that has no exponents higher than 1, and the graph of a linear function is a straight line, whereas a nonlinear function is a function that has at least one exponent higher than 1, and the graph of a nonlinear function is a curved line.

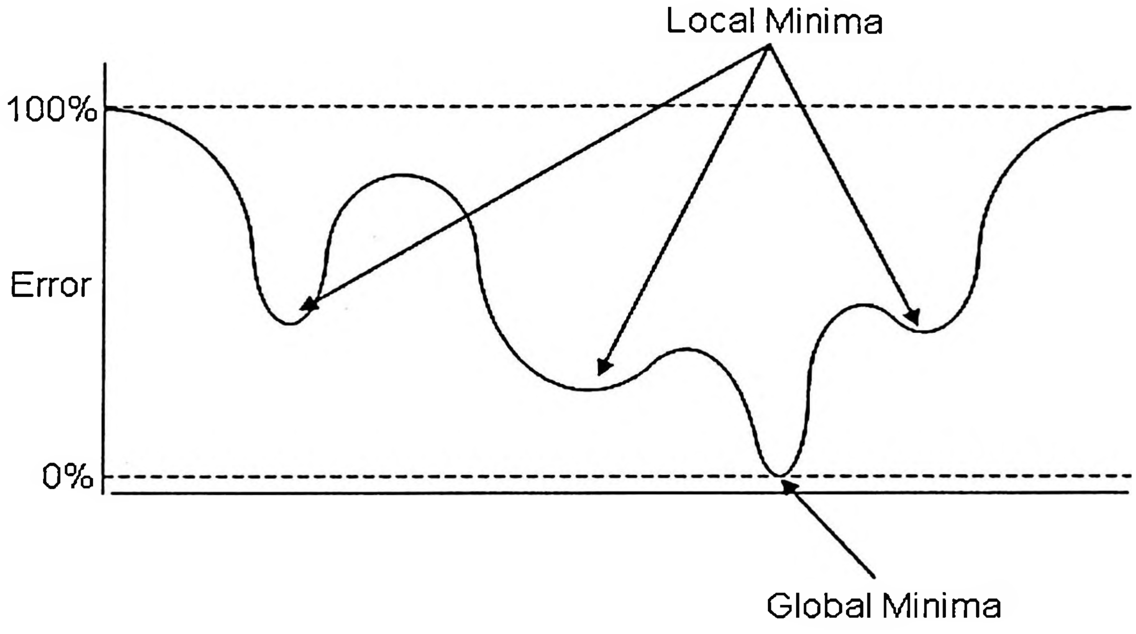

Figure 1.2 Local and global minima (McCullock, 2016)

Local and global minima are terms used to describe solution states with various levels of error. Learning algorithms seek a resting state where error values can be no longer increase, these appear as well as in the wavy line above. A global minima is the ideal solution state when the algorithm rests at a state of zero error, that is when the error is relatively very small nearly to zero per cent. Local minima are the approximate solutions, places where the algorithm rests and the error level is above zero.

The LM algorithm is an iterative technique that locates a local minimum of a multivariate function that is expressed as the sum of squares of several non-linear real-valued functions. It has become the best function as a standard technique for non-linear least-squares problems, and is broadly adopted in various disciplines for dealing with data-fitting applications, parameter estimation, function approximation among others (Sun & Yuan, 2006).

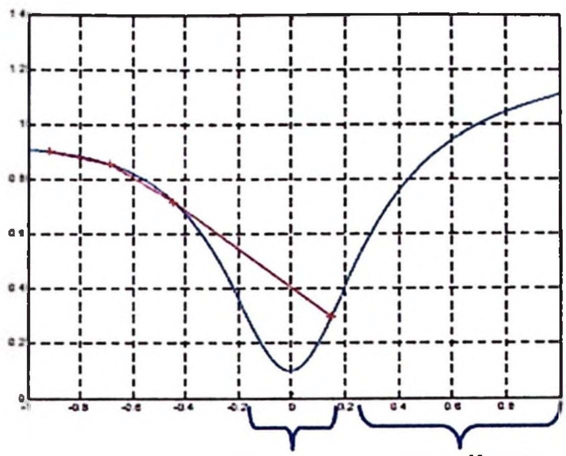

Figure 1.3 The Levenberg-Marquardt method varying between steepest-descent and Gauss-Newton steps (Slobodan, 2016)

Figure 1.3 shows how the Levenberg-Marquardt algorithm interpolates between steepest descent and the Gauss-Newton method. As the current solution is far from a local minimum, the algorithm behaves like a steepest descent method: slow, but guaranteed to converge. That is to say, when the dumping parameter or combination coefficient μ is very large, it acts like the steepest descent. However, when the current solution is close to a local minimum, it becomes a Gauss-Newton method and exhibits fast convergence. That is to say, when μ is very small (nearly zero) it acts like the Gauss-Newton algorithm.

In the LM algorithm, the performance function to be minimised is defined as the sum of squared errors between the target outputs and the network's simulated outputs. The function can be defined as:

$$\min_{x \in R^n} f(x) = \frac{1}{2}\sum_{i=1}^{m}[\mathbf{r}_i(x)]^2 = \frac{1}{2}\|\mathbf{r}(x)\|^2 = \frac{1}{2}\mathbf{r}(x)^T\mathbf{r}(x) \; ; \qquad (1.1)$$

where r: $R^n \to R^m$ is a non linear function of x. If **r(x)** is a linear function, the problem of equation (1.3), $m \geq n$ is the linear least squares problems.

For non-linear least squares problem, it can also be interpreted as solving the system of m non-linear equations, $\mathbf{r}_i(x) = 0, i = 1,2, \dots, m$, where $\mathbf{r}_i(x)$ is called the residual function.

Provided that $\mathbf{f}$ has continuous second order partial derivatives, Taylor series expansion can be written as:

$$\mathbf{f}(x+h) = \mathbf{f}(x) + \mathbf{J}(x)h + o(\| h \|^2) \; ; \tag{1.2}$$

where $\mathbf{J} \in \mathrm{R}^{m \times n}$ is the Jacobian matrix. This is a matrix containing the first order partial derivatives of the function components.

$$(\mathbf{J}(x)_{ij}) = \frac{\partial f_i}{\partial x_j} \; . \tag{1.3}$$

Since, LM algorithm interpolated between steepest descent and Gauss-Newton methods, then review of Gauss-Newton method need to be presented. The Gauss-Newton method also known as the linearization method uses a Taylor series expansion to obtain a linear model that approximates the original nonlinear model and then employs the ordinary least-squares method to estimate the parameters (Jang et al. 1997).

Definition of Gauss-Newton method based on equation (1.1) (Börlin, 2007)

$$\frac{\partial f}{\partial x} = \sum_i r_i \frac{\partial r_i}{\partial x} \; ; \tag{1.4}$$

$$\frac{\partial^2 f}{\partial x^2} = \sum_i \left[r_i \frac{\partial^2 r_i}{\partial x^2} + \left(\frac{\partial r_i}{\partial x} \right) \left(\frac{\partial r_i}{\partial x} \right)^T \right] \; ; \tag{1.5}$$

$$\mathbf{x} = \mathbf{x} - \eta \left(\sum_i \left(\frac{\partial r_i}{\partial x} \right) \left(\frac{\partial r_i}{\partial x} \right)^T \right)^{-1} \frac{\partial f}{\partial x} \; . \tag{1.6}$$

Equation (1.6) is the update rule of Gauss-Newton algorithm by discarded the term $r_i \frac{\partial^2 r_i}{\partial x^2}$ in equation (1.5) Le Roux (2009). The Gauss-Newton method has the following properties: (i) does not require the computation of Hessian matrix (ii) only valid close to the optimum solution, where $r_i = 0$ and (iii) computational cost in $o(d^2)$.

A potential problem with the Gauss-Newton is that $(\mathbf{J}^T\mathbf{J})^{-1}$ might not always exist, as such LM procedure can handle this situation.

Introducing a positive definite matrix $\mathbf{P} = \psi\mathbf{I}$ in equation (1.6), we obtained:

$$\mathbf{x} = \mathbf{x} - \eta\left(\sum_i \left(\frac{\partial r_i}{\partial x}\right)\left(\frac{\partial r_i}{\partial x}\right)^T + \Psi\mathbf{I}\right)^{-1} \frac{\partial f}{\partial x}\ . \tag{1.7}$$

Equation (1.7) corresponds to the LM algorithm, where Ψ is the combination coefficient or learning parameter and $\mathbf{I}$ is the identity matrix.

Summary of the gradient updates between the three gradient-based methods are presented in Table 1.1. The general formula for update rules of the minimisation methods (Slobodan, 2016) is defined as:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \delta x\ ; \tag{1.8}$$

where $\mathbf{x}$ is the parameters vector of the entire training data set, k is the index of iterations, δx is the increment with respect to parameter $\mathbf{x}$.

Table 1.1 Gradient updates

Algorithm	Gradient	Convergence	Computational Complexity
Gradient Descent	$\psi\delta x = -\mathbf{g}$	Stable, Slow	Gradient
Gauss-Newton	$(\mathbf{J}^T\mathbf{J})^{-1}\delta x = -\mathbf{g}$	Unstable, Fast	Jacobian
Levenberg-Marquardt	$(\mathbf{J}^T\mathbf{J} + \psi\mathbf{I})^{-1}\delta x = -\mathbf{g}$	Stable, Fast	Jacobian

Table 1.1 shows the gradient updates for each of the three different optimisation algorithms. The gradients updates formulae are presented in Column 2 of Table 1.1. In Column 3, the convergence for each of the algorithm is described. The Levenberg-Marquardt algorithm provides good convergence of being stable and fast than the other two algorithms. The Gauss-Newton and Levenberg-Marquardt algorithms involve computational complexity for calculation of Jacobian matrix.

1.6.2 The Genetic Algorithm

Genetic algorithm (GA) initially appeared in 1967 in Bagley's thesis "The Behaviour of Adaptive Systems which employ Genetic and Correlative Algorithms" (Bagley, 1967). The theory and applicability of GA was then strongly influenced by (Holland, 1975) who can be considered as the pioneer of genetic algorithms.

The GAs are simulations of evolution, in other words, the GAs are inspired by Darwin's theory of evolution (Chakraborty, 2010). In most cases, genetic algorithms are probabilistic optimization methods that are based on the principles of evolution.

The GAs are numerical search tools aiming at finding the global maximum (or minimum) of a given real objective function of one or more real variables, and subject to various linear or non linear constraints (Marseguerra & Zio, 2000).

The genetic algorithm comprises three phases (operations): mutation, crossover and fitness selection. Mutation models change randomly the genetic information of creatures, and are motivated by random change of genetic information in living organisms. Crossover models the exchange of genetic information of creatures, and is inspired by exchange of genetic information in living organisms. The fitness of an organism is measured by success of the organism in its life (survival) (Schmitt, 2001).

The major advantage of evolutionary algorithms is that they do not have many mathematical requirements on the optimization problems (Bal, Paul, Chakraborty & Sen, 2014).

1.6.3 The Particle Swarm Optimisation

Particle swarm optimization (PSO) is a stochastic optimization technique developed by Eberhart and Kennedy (1995) inspired by the social behaviour of bird flocking or fish school movement.

The PSO is an algorithm modelled on swarm intelligence that finds a solution to an optimization problem in a search space and predicts social behaviour in the presence of objectives. While searching for food, the birds are either scattered or go together before they locate the place where they can find the food. Within their movement for searching food, there is a bird among them that can smell the food very well.

Considering the particle swarm optimization, solution swarm is compared to the bird swarm, the birds' moving from one place to another is equal to the development of the solution swarm, good information is equal to the most optimist solution, and the food resource is equal to the most optimist solution during the whole course (Bai, 2010).

In PSO algorithm, particle swarm consists of "*n*" particles, and the position of each particle stands for the potential solution in D-dimensional space. The particles change its condition according to the following three principles:

- To keep its inertia
- To change the condition according to its most optimist position
- To change the condition according to the swarm's most optimist position.

This method has the advantages of having no mutation calculation and is based on intelligence. However, it has certain disadvantage being in one-sided idealist position, which causes the less exact at the regulation of its speed and the direction.

1.6.4 The Artificial Bee Colony

The Artificial Bee Colony (ABC) is one of the most recently defined algorithms by Karaboga (2005), and is motivated by the intelligent behaviour of honey bees. It only uses common control parameters such as colony size and maximum cycle number.

In the ABC algorithm, the bees in a colony are divided into three groups: employed bees (forager bees), onlooker bees (observer bees) and scouts. For each food source, there is only one employed bee. That is, the number of employed bees is equal to the number of food sources.

The ABC is applicable in many areas including the combinatorial optimization (Pan, Tasgetiren, Suganthan, & Chua, 2011), job scheduling, engineering design, function optimization and the travelling salesman problem. In general, ABC is a global optimization algorithm and employs only three control parameters (population size, maximum cycle number and limit) that are being predetermined by the user.

1.7 Structure of the Thesis

The structure of this thesis includes six major chapters. Each chapter represents an essential part of the main work and is organised as follows:

Chapter 1 provides an outline of the overall research work, which comprises the background of the study, statement of problem, objective of the study, research methodology and an overview of the thesis. Each are described from a point of view to enable thorough understanding of the proposed method with other related existing techniques used.

Chapter 2 presents a review of related literatures using the same data sets as that of the proposed model, conventional ANFIS based Gradient Descent algorithm, Genetic Algorithm, Particle Swarm Optimisation and Artificial Bee Colony.

Chapter 3 presents the theoretical framework of the study. This provides the background that supports investigation obtained from authors of previous research.

Chapter 4 presents in detail the theory of adopted in developing the proposed model.

Chapter 5 focuses on the selected twelve benchmark medical data sets obtained from the University of California at Irvine's (UCI) machine learning repository. The experimental settings of this research are described. The performance of the proposed model is presented and the results obtained are compared with those obtained by using conventional methods and those based on meta-heuristic methods. The results and discussion are presented in this section.

Chapter 6 provides conclusion and recommendations for future research work.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

The fuzzy expert systems can be designed to deal with the uncertainty and imprecision of real world problems. Some components of the system are human-like, adaptable and explanations. Two popular and most powerful soft computing techniques of fuzzy logic and neural networks are combined to produce an integrated fuzzy model that can be applicable in the fields of pattern recognition.

This chapter reviews important areas that are related to this research namely pattern recognition (classification), artificial neural networks (ANN), fuzzy models and neuro-fuzzy models.

Table 2.1 Show summary of different techniques as described in the literatures.

Author	Technique used	Objective
Guvenir and Sirin (1993)	GA-CFP	Classification of 6 different real and artificial data sets.
Pradhan, Korimilli, Satapathy, Pattnaik and Mitra (2009)	SANN	Diagnosis of 4 different diseases.
Ozsen and Gunes (2009)	GA-AWAIS	Determine the weights attributes via genetic algorithm for classification of 2 different data sets.
Li and Lu (2010)	KMP+PSO	Adjusting parameters for classification performance of 4 different data sets.
Vijaya, Nehemiah, Kannan and Bhuvaneswari (2010)	Fuzzy neurogenetic	Predicting the severity of cardiovascular disease.
Chen, Su, Chen and Wang (2011)	PSO+1-NN	Feature selection for diagnosis of medical disease
Karegowda, Manjunath and Jayaram (2011)	GA+BPNN	Classification of Patients with Pima Indians Diabetes disease.
Adam et al., (2012)	ANN+PSO	Prediction performance of an imbalanced data set.

Table 2.1 Continued

Author	Technique used	Objective
Muthukaruppan and Er (2012)	PSO-FES	Diagnosis of Coronary artery disease.
Neshat, Sargolzaei, Toosi and Masoumi (2012)	CBR-PSO	Predict Hepatitis disease.
Palanisamy and Kanmani (2012)	ABC-Boosting	Classification of accuracy with 7 different techniques on 10 different datasets.
Settouti, Saidi and Chikh (2012)	FCM-ANFIS	Diagnosis of Diabetes
Uma and Kirubakaran (2012)	GA-ACO	Diagnosis of Heart disease
Adeli, Bigdeli and Afshar (2013)	GA-ANFIS	Diagnosis of Hepatitis Disease.
Beheshti, Shamsuddin, Ebrahim and Yuhaniz (2013)	CAPSO-MLP	Diagnosis of 9 different medical diseases.
Beloufa and Chikh (2013)	ABC	Diagnosis diabetes disease.
Chiu, Chen, Wang, Chang, Li-Chien (2013)	GFNN+GA	Classification performance of Chronic Kidney disease
Elshazly, Azar, Hassanien and Elkorany (2013)	EI-GA-RS	Classification of medical data sets.
Neshat, Adeli and Masoumi (2013)	GA-CBRPSO	Diagnosis & detection of liver disorder.
Rawat and Burse (2013)	GA-NFL	Predict Ovarian cancer disease.
Schiezaro and Pedrini (2013)	ABC PSO	Data Feature selection with 2 different techniques on Statlog-Heart disease.
Uzer, Yilmaz and Inan (2013)	ABCFS+SVM	Predict Liver disorder disease.
Yigit (2013)	DW-ABC KNN	Classification of 4 different data sets.
Kalaiselvi and Nasira (2014)	ANFIS-KNN	Diagnosis of Diabetes & Liver cancer.
Mangat and Vig (2014)	ACO/PSO	Classification of 8 medical data sets.
Ziasabounchi and Askerzade (2014)	ANFIS	Predict Cleveland-Heart Disease.
Cam, Cimen and Yildirim (2015)	MLP+GA, MLP+PSO, MLP+ABC	Diagnosis of diseases with 3 different techniques.
Sivasankar, Nair and Judy (2015)	PCA+MGA	Diagnosis of 5 medical data sets

2.2 Review of Related Works

Some of the classification methodologies using medical database that were carried out recently are described in this section. Detail on how medical datasets were applied in ANN, ANFIS and other classification techniques are as follows:

Guvenir and Sirin (1993) developed a hybrid system called genetic algorithm based on classification by feature partitioning (GA-CFP). The GA was used to determine the domain dependent feature weights and generalisation limit based on chromosomes for each feature, both of which were real-valued. The standard operators of genetic algorithms, namely, reproduction, crossover and mutation were used. The CFP algorithm was used to train the training examples using the feature weights and generalisation limits, and then tested with the examples in the test set. While in the proposed method, the Modified Levenberg-Marquardt algorithm was used to train the fuzzy inference system parameters and then tested with the examples in the test set. Filter approach using information gain criteria was used for feature selection. The methodology used in the testing of GA-CFP algorithm was the leave-one-out cross validation method, whereas in the proposed algorithm, hold-out validation method was used, because the leave-one-out cross validation method requires more computational capabilities and time consuming than hold-out validation method. For robustness, the existing GA-CFP method used 6 different real and artificial data sets, namely, Iris flowers, Echocardiogram, Cleveland heart disease and Waveform. In the proposed method, 12 different real medical data sets were used.

Pradhan et al. (2009) designed a simple artificial neural network (SANN) model for data classification on Iris, Pima, Blood Transfusion and Ecoli datasets. The Genetic algorithm was used to optimally find out the number of neurons in the single hidden layered model. The proposed model was trained with Backpropagation and Genetic algorithm respectively. However, the designed models were compared with existing model – functional link ANN (FLANN) for data classification accuracies. It was revealed that the proposed models performed better than FLANN model. This existing method was based on ANN approach, while in the proposed method, ANFIS based approach was implemented.

Ozsen and Gunes (2009) developed a system known as genetic algorithms for attribute weighted artificial immune system (GA-AWAIS). The genetic algorithm (GA) was used to determine the weights of attributes. The weights were calculated using statistical information data set, such as standard deviation and mean value of attributes. The existing GA-AWAIS algorithm was applied to 2 different medical diagnosis problems, Statlog Heart Disease and Bupa Liver Disorder Diseases. The procedure used in finding optimum weights via GA are: initialise population of individuals, which represents the weight of features for each class, calculate fitness of these individuals, select a number of individuals having minimum cost values, crossover and mutation rate, and convergence rate. The parameters of GA for Statlog heart disease used were maximum generation number, maxgen = 100, population size, Nind = 50, number of generations used in the convergence test, gen = 10 and mutation rate used in the convergence process, MUTR = 1/(13x2) (13 features for 2 classes). In the proposed method, grid partition method was used to determine the initial values of parameters instead of GA to determine the weight of attributes. Twelve medical data sets were used. Hold-out validation method was used instead of 10-fold cross validation method, because Hold-out validation method do not require knowing the number of parameters or model size, it is also easy to implement.

Li and Lu (2010) described machine learning algorithm known as refining kernel machine pursuit by utilising 3 optimisation methods to perform the refining procedure. The 3 optimisation methods are gradient descent (GD), simulated annealing (SA) and particle swarm optimisation (PSO). Their performances are analysed and evaluated according to experimental results based on 4 different data sets, namely, Heart disease, Pima Indian diabetes, Sona and Ionosphere. The kernel machine pursuit (KMP) was first trained for building a preliminary solution, and then all parameters in the solution were further optimised. The second stage involves testing of the 3 optimisation techniques. Throughout the experiments, 2/3 of examples were randomly selected as training examples, Gaussian kernel functions were chosen for support vector machine in which kernel width and penalty parameters were decided by 10-fold cross validation on training sets. The size of the PSO swarm was set as 20. The best classification accuracy results for Heart, Pima Indian diabetes, Sona and Ionosphere were obtained as 84.33%, 73.07%, 88.30% and 94.59%, respectively. In the proposed method, Sugeno fuzzy inference system was first trained based on grid partition method to determine the initial values of premise parameters, and on the other hand, the final fuzzy model was used to select the best rules for the purpose of adjusting the set of parameters in order to make the performance errors smaller. However, in the proposed method, throughout the experiment, first 2/3 of examples were selected as training examples and the last 1/3 of examples as test examples. Because for an efficient and accuracy of results, the unseen examples should be used for testing of the model.

Vijaya et al. (2010) developed an intelligent model by using a fuzzy neurogenetic approach for predicting the severity of cardiovascular disease. Their proposed method was tested on a Cleveland heart disease dataset from the UCI repository. Trapezoidal membership function for the fuzzification process was used. The number of neurons in the input layer, hidden layer, and output layer were 35, 13 and 5, respectively, in their architecture. Sigmoidal activation function was applied for both hidden layer and output layer neurons. They selected the weight associated between nodes based on the best fitness value of the chromosome by using a genetic algorithm. Their system produced IF...THEN rules with an accuracy of 88.35%. In the proposed method, fuzzy neuro approach with a new learning algorithm Modified Levenberg-Maquardt was used instead of fuzzy neurogenetic approach. Twelve benchmark datasets were used for robustness instead of only one dataset.

Chen et al. (2011) developed an analytical approach by integrating particle swarm optimisation (PSO) to address the problem of feature selection and the 1 nearest neighbourhood (1-NN) method. The method known as PSO+1-NN was implemented based on 8 life science data sets, which helped in identifying important factors and provides a feasible model for diagnosis of medical diseases. The effectiveness of PSO+1-NN algorithm was evaluated and compared with 5 methods, namely, backpropagation neural network (BPNN), logistic regression (LR), support vector machine (SVM), C4.5, decision tree (DT) and integration of genetic algorithm & 1-NN (GA+1-NN). The geometric means (G-Means) was used as the main metric to evaluate the performance of the classifiers. The average classification accuracy for liver disorders was obtained as 68.99%. In the proposed method, filter approach based on information gain criteria was used to address the problem of feature selection instead of 1-NN. The performance error, misclassification rate and performance measuring accuracy, sensitivity and specificity were used for evaluation of the classifier (ANFIS-LMAD) instead of G-Means metric. However, for robustness, twelve benchmarked data sets were used instead of eight data sets.

Karegowda et al. (2011) developed a data mining tool for classification of Pima Indians Diabetes. They presented the application of a hybrid model that incorporates Genetic algorithm (GA) and Backpropagation neural network by using gradient descent method. GA was used for initialization and optimization of connection weights for backpropagation neural network. Two different filter approach namely Decision tree and correlation based feature for identifying relevant features are employed. In their experiment, the best topology used for GA-BPN with four types of crossover and 4 input was 4-8-1 topology, which resulted to 84.71% overall accuracy. In the proposed method, a hybrid model that incorporates ANFIS and ANN learning algorithm is applied. ANN learning algorithm (Modified Levenberg-Marquardt algorithm) was used for the training of premises parameters. This existing method lacks ANFIS based approach, which may produce better accuracy of results than using only ANN based approach.

Adam et al. (2012) described a two-step supervised learning artificial neural network (ANN) for imbalanced data set problems. The method consists of the parameter tuning of ANN's weight based on Levenberg-Marquardt backpropagation learning algorithm, which was utilised in the first step learning. In the second step learning mechanism, optimising the decision threshold of the step function at the outer layer of ANN using particle swarm optimisation (PSO) was introduced. After all the steps learning are accomplished, the best weights and decision threshold value were obtained and used for testing process. Haberman's survival imbalanced data set obtained from University of California at Irvine's machine learning repository was chosen. The prediction performance was assessed by geometric means (G-Means) and was used to indicate the efficiency of the classifier. The classification result for the imbalanced data set was obtained as 70.47%. The proposed method consists of