METAHEURISTIC ALGORITHMS TO ENHANCE ARTIFITIAL NEURAL NETWORK FOR MEDICAL DATA CLASSIFICATION

by

Ihsan Salman Jasim Al Gburi

Electrical and Computer Engineering

Submitted to the Graduate School of Science and Engineering

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

ALTINBAŞ UNIVERSITY

2018

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Doctor of Philosophy.

Assoc. Prof. Khalid Shaker	Prof. Osman N. Ucan
Co-Supervisor	Supervisor

Examining Committee Members (first name belongs to the chairperson of the jury and the second name belongs to supervisor)

Prof. Dr.Osman Nuri Ocan	School of Engineering and Natural Science, Altinbas University	_______________
Assoc. Prof. Dr. Oguz Bayat	School of Engineering and Natural Science, Altinbas University	_______________
Asst. Prof. Dr. Çağatay AYDIN	School of Engineering and Natural Science, Altinbas University	_______________
Asst. Prof. Dr. Adil Deniz Duru	Faculty of sport sciences, Marmara University	_______________
Prof. Dr. Hasan Huseyin BALIK	Air Force Academy, National Defense University	_______________

I certify that this thesis satisfies all the requirements as a thesis for the degree of Doctor of Philosophy.

Asst. Prof. Dr. Çağatay AYDIN

Head of Department

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Ihsan Salman Jasim Al Gburi

[Signature]

Assoc. Prof. Dr. Oğuz BAYAT

Director

Approval Date of Graduate School of
Science and Engineering: ____/____/____

DEDICATION

To the soul of my father who passed away before few months.

To the great lady my mother who always kept praying for me.

To my family my wife and children

To my brothers and sisters

ACKNOWLEDGEMENTS

All the praises be to the mighty Allah, the Merciful and the Beneficent for the strength and blessing in the completion of this study.

Indeed, there are many wonderful people who have contributed significantly throughout the whole course of my study up to the completion of this thesis. I owe a great deal to them.

First and foremost, I wish to express my most sincere acknowledgment to my supervisor: Professor Dr. Osman Nuri Ucan for his valuable guidance, generosity and freedom throughout the entire research and thesis writing. Sincere appreciation goes to co-supervisor Assoc. Prof. Dr. Khalid Shaker for his support and constructive comment. Many thanks for our graduate school director Assoc. Prof. Dr. Oğuz BAYAT for the encouragement, thoughtful comments and helpful discussion.

ABSTRACT

METAHEURISTIC ALGORITHMS TO ENHANCE ARTIFITIAL NEURAL NETWORK FOR MEDICAL DATA CLASSIFICATION

[Al Gburi, Ihsan],

PhD, Graduated School of Science and Engineering, Altınbaş University,

Supervisor: Professor Dr. Osman Nuri Ucan

Co-Supervisor: Assoc. Prof. Dr. Khalid Shaker

Date: [September 2018]

The tremendous growth of computer hardware technologies and their abilities to solve huge amounts of complex of data has motivated researchers to overcome complicated data mining challenges and problems. Medical dataset classification represents one of the most crucial and complicated problems faced by researches in the field of artificial intelligence and data mining. The different diseases and the various ways of diagnosis by using multiple testing have produced large amounts of complex medical data. Moreover, the huge number of patient records in clinical centers and hospitals and other health institutions has generated the need for advanced and accurate medical mining applications to help doctors and therapists investigate cases regardless whether patients are in critical conditions or require remote follow-ups.

This thesis focuses on the hybridization of the artificial neural network (ANN) and metaheuristic algorithms to enhance the accuracy of a classification model for the overlapping fields of medical data mining. The key problems associated with medical diagnoses involve the identification of highly accurate classification models. The contributions of this thesis revolve around the two important classification problems or issues highlighted in the related literature. For the first strategy, the relation between the ANN structure and the optimized algorithm is established. For the second strategy, the tradeoff between diversification and intensification is investigated as part of the search for the optimal global solution.

The first part of this paper discusses the effect of metaheuristic iteration on ANN structure. Here, ANN is enhanced using three metaheuristic algorithms (PSO, GA, and FW). The proposed models are tested on five standard medical benchmarks. The study proposal is successfully implemented, and remarkable results are obtained.

The second part of this paper investigates the tradeoff between exploration and exploitation when obtaining the optimal global solution. The number of hidden layers and the number of neurons in each layer can both affect ANN learning. Thus, the ANN used in this thesis involves the selection of a complex structure that can achieve highly accurate results; consequently, metaheuristic algorithm efficiency can be guaranteed when searching for the global optimum. Two metaheuristic algorithms are combined to formulate a new and improved algorithm for any problem domain. However, selecting the highly accurate two algorithms is not mandatory; instead, empirical tests can be performed for convenience.

Keywords: Medical Information Systems, Diseases Diagnoses Systems, Optimization Algorithms

TABLE OF CONTENTS

Pages

LIST OF TABLES

Pages

LIST OF FIGURES

Pages

LIST OF ABBREVIATIONS

ANN	Artificial neural network
GA	Genetic algorithm
PSO	Particle swarm optimization
FW	Fireworks algorithm
LD	Liver Disorders
PID	Pima Indian Diabetes
WDBC	Wisconsin Diagnostic Breast Cancer
NFLT	no-free-lunch-theorem
PD	Parkinson’s disease
DE	Differential evolutionary algorithm
KNN	k Nearest Neighbor algorithm
SVM	Support vector machine
SA	Simulated annealing
TS	Tabu search

1. INTRODUCTION

Data mining is the study of observational datasets to discover relations and to summarize the data in behavior that are both understandable and useful to human. The idea of data mining is not totally new. People have been looking for patterns in data since human life started: farmers required patterns in crop growth, Hunters have been seeking patterns in animal migration activities, politicians seek patterns in elector opinion, and lovers try to find patterns in their partners' responses as shown by Chakrabarti et al. [1]. The primary book in data mining came out in 1992 by Frawley et al. [2].

In recent years data mining has involved great interest in the information systems. This is because current computers are able to create and stock up almost unlimited datasets. In reality, database and information technology has been growing methodically from primal file processing systems to complicated and powerful database systems. The fast growth of database system is illustrated in Figure 1.1.

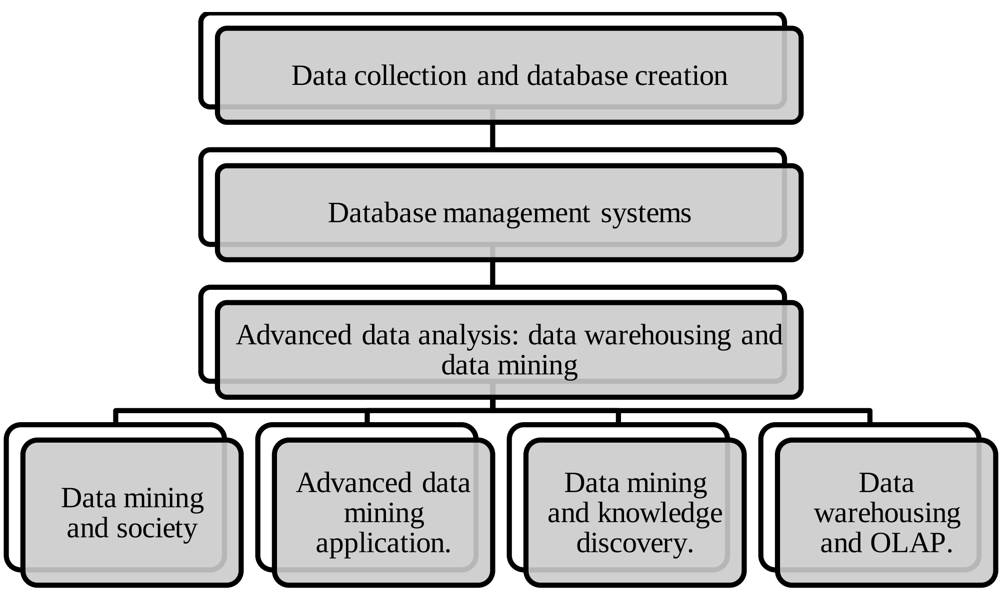


Figure 1.1: The evolution of database system technology.

However, there is no brief set of factors that can completely explain the condition of real-world and complex systems. In recent years, this expansion was continued using a more complicated and intelligent process to achieve the goal of any data mining tasks. The computer scientists are motivated toward development of advanced and intelligent data mining which may utilize methods such as Artificial Neural Networks (ANNs), Genetic Algorithms (GAs), or any other heuristic algorithms as in Han et al. [3]. In fact, data mining adopts techniques from many domains and is an inter-disciplinary field using disciplines such as artificial intelligence, machine learning, information science, statistics and other areas as shown in Figure 1.2.

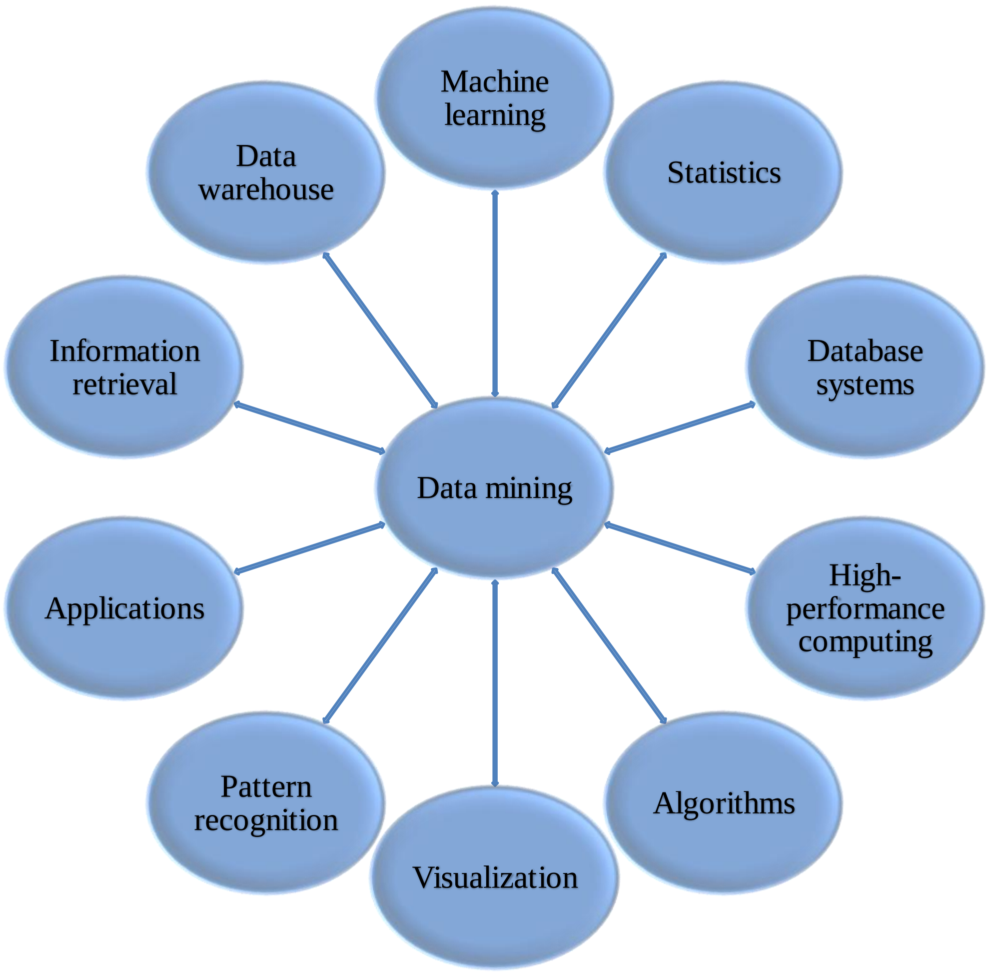


Figure 1.2: Data mining as a confluence of multiple domains

However, each regulation has its own properties that make it useful for certain categories of problems. Statistical techniques are not suitable to address more challenging problems in data mining, especially when there are large datasets Han et al. [4]. Intelligent systems which are all about learning rules and patterns from the observed data provide a great potential for advanced and theoretical research in data mining area.

An ANN is an imitation of a biological natural neural network. ANNs try to generate machines that act in a similar method to the human brain by structuring those using mechanisms that perform like biological neurons. However, the process of ANNs are far more simplified that the procedure of the human brain.

Each ANN is structured into several interconnections of simple processing units which are called nodes or neurons as mentioned in Fausett [5]. An ANN can have one or more layers of nodes between the input layer and output layer. Each of these interlayers are called a hidden layer, and they can be completely or partially connected Fausett [5]. Each connection between two nodes has a specified weight.

Classification consider an important task in data mining area. This thesis focuses on improvement of the ANN models by development hybridization between ANN and metaheuristic approaches and ANN with hybrid metaheuristic approaches to enhance medical data classification accuracy.

1.1 SIGNIFICANT OF THE THESIS

Medical diagnosis is considered as a very significant task that requires being properly and sufficiently executed, thus, designing accurate systems in this area would be very beneficial in diseases diagnoses and physicians aid in diagnosing very complicate and complex cases with huge number of patients in remarkable time with predefined diagnosing systems based on accurate models. An automatic medical diagnosing system would probably be extremely advantageous by bringing all materials, tools, and objectives together in order to achieve diagnose target.

Medical data analysis and diseases diagnosing as acknowledged discovery are important but yet hard errands and naturally are based on years of practice of a professional as describe in Saez et al. [6]

Initiate medical diagnosis model from patients’ medical reports in early time has an important meaning for accurate health treatment to human. This consider the idea behind diseases diagnosis models as X. Liu et al [7].

1.2 PROBLEM STATEMENT

Classify medical data sets studied and performed with different machine learning methods like neural network (NN) as in Kayaer, K., & Yildirim, T. [8], support vector machine (SVM), decision tree (DT) … etc.) all mentions at Pham and Triantaphyllou [9], main objective from these studies was to reach better accuracy performance with less error as much as possible.

Classification methods enhanced by optimization algorithms and techniques to overcomes limitations like overfitting overgeneralization balance in Pham and Triantaphyllou [9], and in Pham and Triantaphyllou [10], cost of using all attributes (dataset features) proved usefulness and worthless where feature reduction methods proposed as in Ahmet Mert et al. [11], or extract extra features as in Zeng et al. [12], all that to reach better evaluate methods (hybrid methods) for mining diseases and search for optimal solution in search space further discussions about metaheuristic in Blum [13].

Due to the learning ability and nonlinearity method of ANNs, anywhere that there are problem of classification or prediction, ANNs are being established. In the literature, many algorithms have been applied to tune the weights of the ANNs to train the ANN model as in Sanggil & Isik [14]. In this case, the most significant factor in the success of an ANN is the selection of the structure of ANN, which necessitates the determination of the proper number of layers and the number of nodes in each layer. ANN structure must perfectly fit the underlying relationship expressed by the observed data as in Yang & Chen [15].

A structure that is too large may cause over fitting. Conversely, a small structure saves computational time but may not accurately fit the observed data. Thus, structure selection must consider both the complexity of the network and the accuracy of fit. Many researchers with applying self-organizing neural network attempt to optimize the structure of neural network and they did not use any algorithm for weight optimization (Oh & Pedrycz [16]; Oh et al. [17]; [18]). This may lead to poor efficiency of the neural network model. Therefore, optimization of the weights and structure of ANN using a capable meta-heuristic algorithm to build an efficient neural network model is considered as research potential. Toward this aim, an efficient solution representation is required to optimize the weights and structure of the neural network instead of optimizing the weights or structure in isolate. Only a few number of researchers attempt to optimize the weights and structure of neural network (Ludermir et al. [19]; Yang & Chen [15]; Zanchettin et al. [20]). There is a solution representation for this problem in the literature (Ludermir et al. [19]; Zanchettin et al. [20]) that a maximum number of connections is defined first, then with 1 and 0 the activity or inactivity of the weights (connections) are determined. Another approach (Yang & Chen [15]) considers that the connection is not included in the structure of neural network if the associated weight value is equal to 0. The problem with these solution representations is time consuming process due to requirement of checking the entire length of solution to identify whether it is active or inactive.

Furthermore, selecting parameters of the algorithm is an important issue in the area of the optimization (Bo & Gallagher [21]; Eiben & Smit [22]; Yuan & Gallagher [23]). Most of the algorithms that are available in the literature have some parameters that need to be tuned in advance. Parameter tuning can improve the performance of the algorithm when solving any problem (Mousavi et al. [24]; Smit [25]). Parameter tuning can be distinct as the problem of searching through all potential parameters for a given problem solving. However, in practice, it is not known in prior which parameter values should be used and finding the best value is challenging.

For the optimization strategy a capable algorithm is required to lead the optimization process to the global optimum solution (Słowik [26]). A balance of exploration and exploitation strategy in search space helps to achieve the global optimum (al-Naqi et al. [27]; Valizadegan et al. [28]). Exploration (diversification) is the ability to test various regions in the problem space to locate a good optimum, hopefully the global one. Exploitation (intensification) is the ability to concentrate the search around a promising candidate solution to locate the optimum accurately. There is a natural trade-off between high exploration and good exploitation (Crepinsek et al. [29]; Tokic & Palm [30]). However, the search for a proper exploration and exploitation trade-off remains as a challenging task in any optimization process.

Moreover, any optimization algorithm searches for a solution that falls in the optimality region. The probability of finding the solution in the optimality area is basically the volume of the optimality area divided by the volume of the search space. This probability is decreased exponentially when the volume of the search space increases (Binneng et al. [31]; Tan et al. [32]). The maintenance of diversity of the solutions in the population is required to ensure that the search space is effectively searched. Searching through the search space with lack of variation is considered as the key reason for early convergence (Niu et al. [33]; Niu et al. [34]). Given this explanation, maintaining the diversity of the solutions in the population with an efficient strategy is needed.

There are no superior metaheuristic algorithms work with all optimization problems. According to no-free-lunch-theorem NFLT there is no universe metaheuristic approach overcome all other methods to optimize all problems (Wolpert and Macready [35]; Ho and Pepyne [36]). NFLT depicted in multiple optimization areas (Aljarah et al. [37]). Principle of NFLT found in medical mining problem (Salman et al. [38]).

Applicability of the model for any organization in the real-world problem is another issue in the area of the research. Constructing the link between research and practice is important subject in the real-world and has received increased attentions (Kazdin [39]; Atkins et al. [40]). To make this link possible, an investigation of applicability of the research is required.

Along literature survey of diseases knowledge discovery or (diseases data mining), a lot of classification methods have been tested different techniques used, and promising results been gathered. But researches in (diseases data mining) is still an open field that for several reasons like:

- Not all the diseases tests learned in disease mining.
- There are always new diseases and new tests could discover for famous diseases.
- Most of diseases classification did not reach ultimate result (i.e. classify methods with 100% accuracy).
- Scientists always develop new mathematical formulas machine learning methods, optimization algorithms, which hybridize and modified to give better solutions.

In this work ANN medical classification methods will develop with aid of metaheuristic optimization algorithms and strategies to enhance ANN diseases classification.

1.3 RESEARCH QUESTIONS

Medical data mining classification problems are like other classification methods from different areas they could be done by classification methods but as a problems data sets medical data sets have specified characteristics that's it came from medical analysis and tests so they are normally big complex finite discreate data. According to this medical data mining classification attend to apply same classification models but in specific style of method setting.

The following research questions are prepared based on previous explanation of the problem of optimization of ANN with metaheuristic algorithms.

1. Does optimization of the weights with changing structure of the ANN build a more accurate model compared to only optimization of the weights or structure in isolate?
2. Does number of iterations of metaheuristic approaches affect to find suitable ANN set of weights for accurate ANN model?
3. What is a proper solution representation that leads the optimization process to a capable model?
4. Do parameters tuning for metaheuristic algorithms may enhance the ability of the algorithm?
5. How to balance the exploration and exploitation strategy of the algorithm to improve the quality of the search process?
6. How hybridization between two metaheuristics may assist ANN?

1.4 RESEARCH OBJECTIVES

The main aim of this research is to investigate the performance of the applying metaheuristic techniques in order to obtain optimal weights of the neural network to have a capable model for diseases diagnosing and classification problems. To achieve this major aim, two objectives are outlined:

1. To determine a suitable solution representation that can optimize the weights with different neural network structure using different number of metaheuristic iterations.

2. To hybridize between evolutionary and trajectory metaheuristic method and trade-off between exploration and exploitation using modified differential evaluation algorithm DEA hybridized with simulating annealing to enhance ANN to classify Anemia new real data set.

A summary of mapping between research questions, research objectives and the contributions is provided in Table 1.1.

Table 1.1: A summary of mapping between research questions, research objectives and contributions in this study

Chapter	Research question	Research objective	Contribution
Third chapter	• Does optimization of the weights with changing structure of the ANN build a more accurate model compared to only optimization of the weights or structure in isolate? • Does number of iterations of metaheuristic approaches affect to find suitable ANN set of weights for accurate ANN model? • What is a proper solution representation that leads the optimization process to a capable model?	• To determine a suitable solution representation that can optimize the weights with different neural network structure using different number of metaheuristic iterations.	Impact of metaheuristic iteration on ANN structure
Forth chapter	• How to balance the exploration and exploitation strategy of the algorithm to improve the quality of the search process?	• To hybridize between evolutionary and trajectory metaheuristic method and trade-off between exploration and exploitation using modified	Improving anemia classification by evolutionary trajectory metaheuristic hybridization.

	• How hybridization between two metaheuristics may assist ANN? • Do parameters tuning for metaheuristic algorithms may enhance the ability of the algorithm?	differential evaluation algorithm DEA hybridized with simulating annealing to enhance ANN to classify Anemia new real data set.	

The first three research questions are answered by the first objective of this research. To achieve this research objective three metaheuristic algorithms used with different number of iterations and tuning their parameters to enhance ANN with more than one structure and the impact of this proposed methodology is investigated in this research.

Research questions number five and six answered by second research objective that is achieved by improve anemia ANN classification by modified hybridization between evolutionary and trajectory metaheuristic algorithms.

Answer of the forth research question could be found with both contributions and included with all research objectives.

1.5 RESEARCH SCOPE

This research focuses on study about ANN model for solving medical classification problems. The main purpose is to improve the quality of the ANN model to have better classification accuracy and forecasting with using metaheuristic approaches. In this research different enhancement techniques are investigated in order to improve ANN structure for accurate medical data classification model.

To test the proposed method the standard accuracy measure used to evaluate classification method when compare models or compare with the literature, different ANN models enhanced by metaheuristic algorithms and hybridized metaheuristic models to improve diversification and intensification for better search space exploration and intensive local exploitation. model tested on real world disease anemia medical data sets and a comparison has been made to find better model for this data set classification.

Statistical tests conducted on every contribution proposed in the context of this study to show the significance of the proposed model, further discussion in section **Error! Reference source not found.**.

1.6 OVERVIEW OF THE THESIS

This thesis consists of five chapters. Chapter I presents the background, problem statement, objectives and aims of the research. The remainder of this thesis is organized in the following way:

Chapter II introduces the problems description and summarizes the approaches applied on the problem. It reviews the history of the methods used for classifications and metaheuristic algorithms and discusses about the techniques that can improve their performance. Then reviews the current techniques applied for other algorithms in the literature.

Chapter III proposes a unify the structural change in ANN and the number of iterations of metaheuristic algorithms. The effect of the selection of different numbers of iterations for more than one structure of ANN is accordingly studied. Proposed methodology tested with big data classification that is contains thousands of attributes. The results of the methods are compared to other techniques in the literature for classification. Finally, the significant of ANN classification and ANN enhanced by metaheuristic algorithm classification statistically tests are discussed.

Chapter IV presents the details of hybridization between Differential evolution (DE) and simulating annealing (SA) to enhance artificial neural network and proposed ANN_DESA method in order to balance between exploration (diversification) and explanation (intensification), the method tested on Anemia real medical data set classification. Anemia data set obtained from (Al-Anbar health directorate/IRAQ). This data was collected from Anbar province hospital laboratories from patients have anemia disease. Proposed methodology begins with choosing best ANN structure (best number of hidden layers and best number of neurons in each layer) then classification process tackle with hybridize DEA and SA to obtain batter classification accuracy. Statistical t test of ANN against ANN+DEASA performed to test the statistical difference significant, p value of null hypotheses obtained to find difference significant.

Chapter V presents the summary of the work, conclusion and future work.

2. LITERATURE SURVEY

2.1 INTRODUCTIONON

This chapter provides an overview of medical classification problems and a review of the techniques for solving these problems that have been proposed in the literature to date. It especially focuses on examining a variety of artificial neural network models to identify an appropriate technique for classification problems.

Following sections of this chapter are arranged as follows: Section 2.2 provides a brief description of classification problems. Extensive review of techniques applied on medical data classification is presented in Section 2.3. An overview of findings from literature review is given in Section 2.4.

2.2 CLASSIFICATION PROBLEM DESCRIPTION

Data mining techniques consist of a set of algorithmic methods to extract the relationships in data (Ngai et al. 2011 [41]). These techniques differ in terms of the problems that they are able to solve (Giraud-Carrier & Povel 2003 [42]). These problems include association, classification, clustering, prediction, outlier detection, regression, and visualization. The review of the literature presented in this paper focuses on methods to solve medical data classification problems.

Classification is the procedure of inference a model (or function) that defines and differentiates data classes or concepts Han et al. 2006 [3]. One of the purposes of data mining is to place raw data into one of several predetermined classes. Scientifically, a class Ci (i^{th} Class) is distinguished as Eq. (2.1) as follows:

$$C\{x \in S \mid f(x)\} \tag{2.1}$$

where object x is mapped from the training dataset S after assessment of the condition function f(x) for x a member of the class Ci. Simply put, classification concerns developing a model to predict the categorical names of unknown objects to differentiate between instances of different classes. This is a basic problem in data mining and is essentially a data analysis task, where a classifier is constructed to predict categorical names. These categories can be presented by discrete values such as 1, 2 and so on. The term ‘objects’ refer to packed data units that are specific to an exact problem, which is known as a pattern.

Classification methods are broadly categorized into supervised (Beniwal & Arora [43]; Caruana & Niculescu-Mizil [44]; Kotsiantis [45]) and unsupervised (Deepthi et al. [46]; Sathya & Abraham [47]) approaches.

In supervised approaches the rules learned during the training process are used for prediction. Rules normally consist of a group of conditional attributes and a class attribute as a decision attribute. The rules are used to classify unseen instances into different classes.

Data classification is a two-step process. In the first step, a classifier describes a determined set of data concepts. This is the training phase, where a classification algorithm learns from the training set. Each instance in the training set is considered to belong to a predetermined class. The classifier is built based on the results of the training process. In this first step, a plan to be developed that divides the data classes. Typically, this is described in the form of classification rules, decision trees, or mathematical formulas, but normally it is in the form of classification rules.

In the second step, the classifier is used in the classification phase when the predictive accuracy of the classifier is measured. If a training set is used to evaluate the accuracy of the classifier, this evaluation is often optimistic because the classifier watches over unsuited data. The accuracy of the classification of a test set of data is based on the percentage of test set tuples that are properly classified by the classifier.

The whole of the above process begins with collection of proof obtained from various data sources. Ideally, the data should be of small size, independent and discriminative. Raw data does not easily satisfy these criteria and therefore a set of procedures such as feature generation, extraction and selection is needed to produce a relevant input for the classification procedure.

The building of the classification model consists of four phases:

- Technique selection, which can be difficult if there is no strong knowledge about the character of the problem.

- Data preprocessing, which is used to enhance the prediction ability of the model.

- A learning process using training data, which as mentioned above can be in the form of supervised or unsupervised learning.

- A testing process that evaluates the model built in the learning phase by using it on test data.

While Kalantari et al. [48] illustrate classification process with four steps or phases: 1) Preprocessing phase, 2) Learning phase, 3) Performance evaluation phase, and 4) Decision phase as shown in

Figure 2.1 .

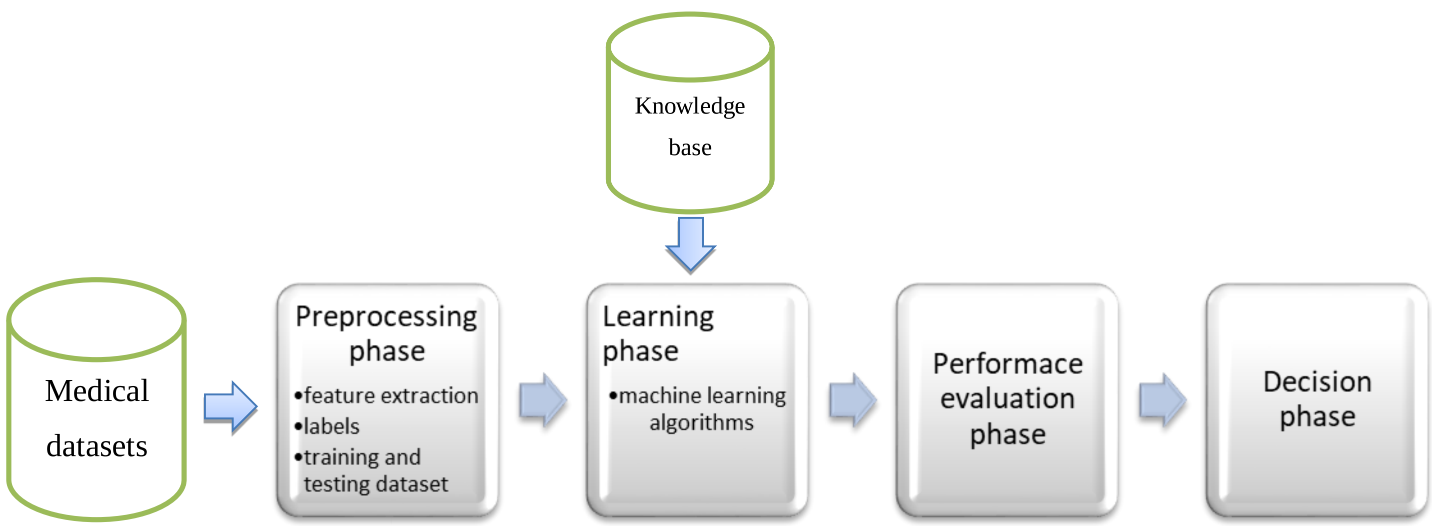


Figure 2.1: classification learning system

2.3 APPROACHES APPLIED TO MEDICAL PROBLEMS

In medical data mining Classification consider a prediction methods or models to find categorical classes like ‘yes’, ‘no’ or ‘0’,’1’, …; or predict a value of discrete values Avci & Dogantekin [48], for example classify medical data to discover the type of treatment, is it type 1, type 2, or type 3.

In this study we will call first type binary class classification and will call second one multi class classification to differentiate between the two processes during implementation.

In this review the most recent methods in the literature are highlighted and summarized to provide an overview of the techniques currently available. first the single techniques applied to the classification problem are discussed and then the hybridization metaheuristic enhancement are studied and are examined through state-of-the-arts. A summary of the methods applied in classification and related works is shown in Figure 2.2.

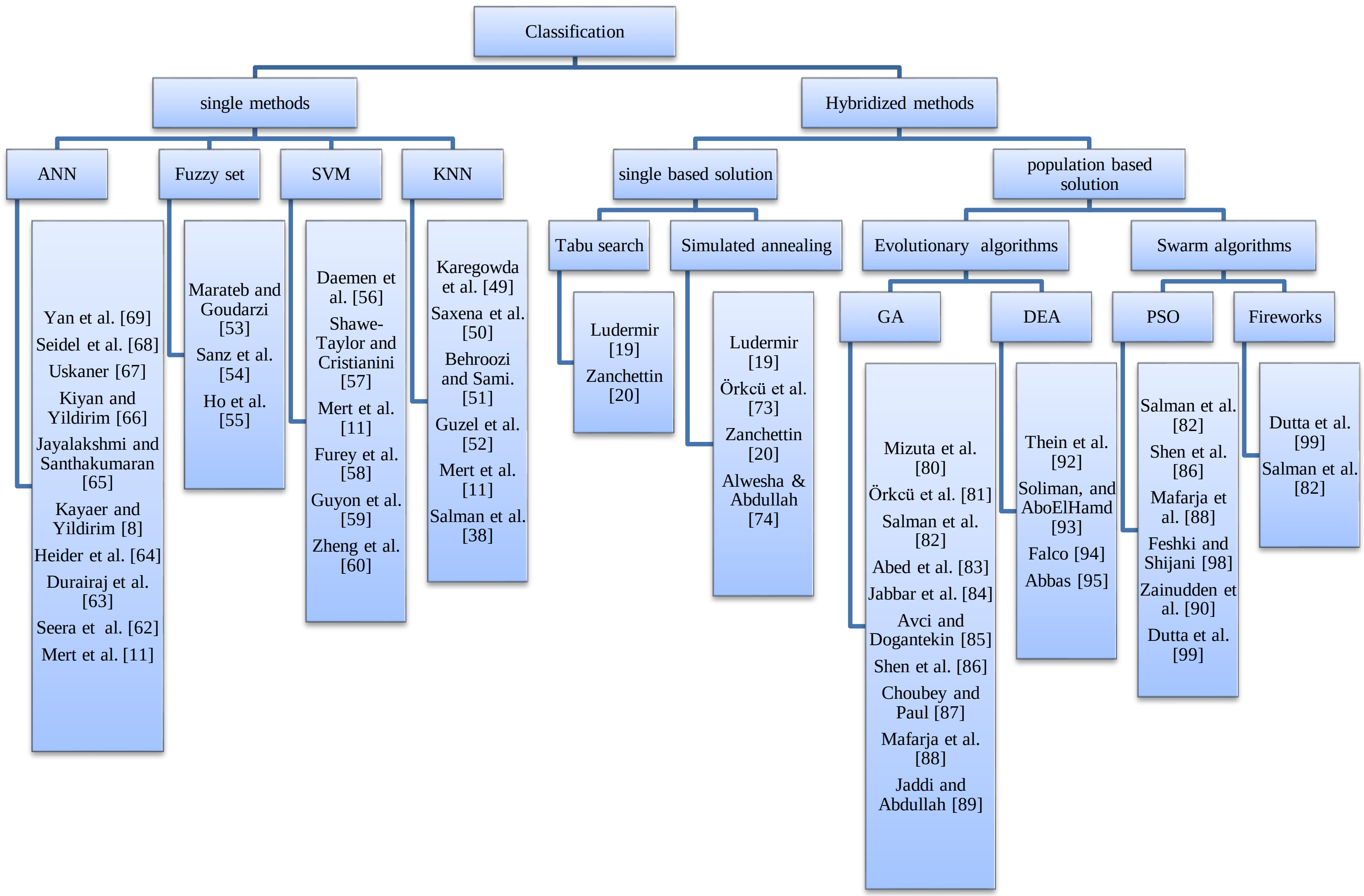


Figure 2.2: A summary of methods applied for classification problem

2.3.1 Classification approaches

There are number of machine learning models used for classification, in the context of medical data mining, classification techniques are mainly categorized into the following main different types (but not restricted to) according to the literature review: lazy learner methods, fuzzy set, kernel, and artificial neural network.

2.3.1.1 Lazy learner K-nearest neighbor KNN

According to Han & Kamber [3], lazy learner consider a type of machine learning methods that is the learning process starts when new tuple or the test tuples starts, and there is no model build during training phase, in the other hand eager learners used training tuples or samples to formulate model and used this model to predicts classes for new test tuples.

KNN, K-nearest neighbor first introduced in early 1950s, but due to computation limitations KNN stay idle until the increasing of computing power at 1960s the methods became popular and deployed widely in different fields.

K-nearest neighbor classification algorithm KNN is supervised classification algorithm depends on the distances between the test dataset and the training dataset and finds out which one is closest and take the majority class from K-list according to K samples chooses randomly.

Distance could be found in different ways mostly used Euclidian's distance shown

$$d(p_i, q_i) = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2} \tag{2.2}$$

This algorithm normally used when there is online new instance appears or new set of instances, and need to be classify according the previous data with output class.

In Karegowda et al. [49], categorized diabetes patients by cascading K-mean clustering and K-nearest neighborhood classification and perform classification within 3 stages, first, K-mean, second, Correlation based feature selection CFS, Finally, KNN.

Saxena et al. [50] , classify diabetes in mellitus dataset with K-nearest neighborhood.

Pima Indian Diabetes classified in spiral spinning technique with KNN and ANN in Salman et al. [38].

Behroozi and Sami. [51] proposed KNN classifier with other multi-classifiers to discriminating Parkinson's disease. Methodology was to use different methods for classification the final classification result would be the majority vote from all the classifiers.

Guzel et al. [52] develop approach to assign missing values with KNN and Naïve Bayes. Then, the performance of the system is evaluated by KNN and Naïve Bayes classifiers to detect breast cancer.

Mert et al. [11] explore reducing problem dimensionality by using independent component analysis ICA with KNN, ANN, SVM, and RBFNN classifiers. Models tested with different cross validation. Proposed study uses Wisconsin diagnostic breast cancer as problem domain with thirty features.

2.3.1.2 Fuzzy sets

Fuzzy classification is the procedure of combining elements into a fuzzy set. In this approach, the membership function is distinct by the truth value of a fuzzy propositional function Zadeh 1965 [53]. According to this method elements consists of many logics valued in which the truth values of variables might be any real number between 0 and 1 considered to be fuzzy. Elements of fuzzy sets have degree of membership.

In term of medical applications fuzzy sets classification methods used in many researches.

Marateb and Goudarzi [54] proposed a fuzzy rule base system to diagnose Coronary heart diseases and coronary artery diseases, the proposed method show agreement with the standards.

Sanz et al. [55] proposed new methodology to diagnose cardiovascular disease within the next 10 years by association fuzzy rule-based classification systems with interval-valued fuzzy sets to build a classifier able to deal with the issues of identifying the risks for a patient that suffer from this disease.

Ho et al. [56] propose an interpretable gene expression classifier iGEC with fuzzy rule base for data analysis. Proposed iGEC has three aims to be instantaneously optimized: maximum accuracy, minimum rule numbers, and minimum genes.

2.3.1.3 Kernel methods SVM

Another supervised machine learning technique is the support vector machine (SVM), this method is one of the best kernel methods which has powerful class of algorithms for pattern analysis as mention in Daemen et al. [57] and Shawe-Taylor and Cristianini [58]. This technique has been used in many research studies on the classification problem.

Furey et al. [59] developed a new method using SVMs. Their analysis focused on the classification of tissue samples and an investigation of the data for mislabeled tissue results. Guyon et al. [60] addressed the problem of assortment of a small subset of genes from gene expression data recorded on DNA micro-arrays by using available training samples from cancer and non-cancer patients.

Mert et al. [11] explore reducing problem dimensionality by using independent component analysis ICA with KNN, ANN, SVM, and RBFNN classifiers. Models tested with different cross validation. Proposed study use Wisconsin diagnostic breast cancer as problem domain with thirty features.

Zheng et al. [61] proposed breast cancer classification system uses a hybrid of K-means and support vector machine to extract useful information. K-mean used to extract 6 features from 32 features of (WDBC) data set, then (SVM) is used to obtain the new classifier to differentiate the incoming tumors.

2.3.1.4 Artificial neural network

Artificial neural networks are widely used for classification and rule generation. Among the vast number of ANN methods that have been used for classification, the simplest is that based on a fixed number of nodes and layers. These methods were initially based on the ‘backpropagation’ neural network (Burks et al. 2000; Gudise & Venayagamoorthy 2003; Güler & Derya Übeyli 2003; Horikawa et al. 1992; Riedmiller & Braun 1993). Backpropagation is a shortened version of the term ‘backward propagation of errors’. In this method, the ANN has to be taught by using a training dataset. The training dataset consists of input values allocated with a matching goal (desired output). The training of the network is performed by an iterative procedure. In each iteration, weights are adapted using new data from the training dataset. Each teaching step begins with forcing input signals from the training set. After this step, the output values for each neuron in each layer can be determined. Then, in the next step, the output value of the network is compared with the desired output value, which is found in the training dataset. The difference between the network output and desired value is called the error value.

The idea of backpropagation is to propagate the error value back to all neurons into which output values were input for the examined neuron. The weights used to propagate errors back are the same as those used through computing the output value. It is just the direction of the data flow that is altered.

Mert et al. [11] explore reducing problem dimensionality by using independent component analysis ICA with KNN, ANN, SVM, and RBFNN classifiers. Models tested with different cross validation. Proposed study use Wisconsin diagnostic breast cancer as problem domain with thirty features.

Seera et al. [62] proposed a hybrid intelligent system that consists of fuzzy min-mx neural network as Classification and Regression Tree, and the Random Forest model, this hybrid system aims to exploit the advantages of the constituent models and, at the same time, alleviate their limitations, able to learn incrementally using Fuzzy Min–Max neural network, explain its predicted outputs with the Classification and Regression Tree and achieve high classification performances by Random Forest.

Durairaj et al. [63] diagnose Pima Diabetes data set with back propagation Network which was trained by Levenberg–Marquardt (LM) algorithm. The study observed that Neural Network structures could be effectively used to find better results. The classification accuracy of BPN with LM obtained by this study was better than those get by other studies for the predictable validation model.

Heider et al. [64] introduced a neural network cluster that included four subfamily networks to designate a small GTPase to a subfamily and a filter network to classify small GTPases.

Kayaer and Yildirim [8] diagnose Pima Indian Diabetes PID with three ANN models multilayer perceptron MLP, radial basis function RBF, and general regression neural network GRNN. The study tests the models for competition in term of accuracy.

Jayalakshmi and Santhakumaran [65] reveal PID ANN classification with different normalization preprocessing techniques in order to control neural model complexity, proposal prove in term of chosen medical dataset statistical column normalization preprocessing give a batter accuracy then the others.

Kiyan and Yildirim [66] classify Wisconsin breast cancer data WBCD with four ANN models statistical neural network structures, radial basis network RBF, general regression neural network GRNN and probabilistic neural network PNN. The study tests the models for competition in term of accuracy.

Uskaner [67] diagnose Haberman Survival dataset with multi-layer perceptron neural network by tuning MLP parameters and show the effect of choose different values on the model error.

In Seidel et al. [68] MLP employed to determine cancer from urinary nucleosides analysis the results validate the utility of ANN method comparing with existing studies.

Heart disease medical dataset diagnoses by MLP depending on decision system in Yan et al. [69], the performance of classification assess by three performance assessment.

All classification techniques have strengths and weaknesses the relative importance of which depends on data being analyzed. In this literature review the strengths and weaknesses of each technique in respect of solving the classification problem are provided to highlight the problems, limitations, and gaps in the proposed methods. Moreover, the strength and efficacy of these methods are also investigated so that their respective advantages could be leveraged in further studies. An in-depth understanding of these methods’ strengths and limitations would enable the appropriate improvement of existing techniques in order to develop an advanced and intelligent method for solving the classification problem. The principal strengths and limitations of the various classification methods reported in the literature are summarized in Table 2.1

Table 2.1: Strengths and limitations of the classification methods

Classification method	Strength	Limitation
KNN	• there is no model build during training phase	• The learning process starts when new tuple or the test tuples starts • The method classifies when new instant arrives according the previous data with output class • There is no statistical or mathematical method to determine best K value.
SVM	• Support vector machine can be easily extended to perform numerical calculations. • Support vector machine is very useful for general pattern recognition, regression and classification.	• Lack of clarity in the results • Support vector machine is computational inefficiency to finding an approximate minimum enclosing ball a set of instances.
Fuzzy set	• Ability to examine each fuzzy if-then rule.	• Exponential increase in the number of possible fuzzy if-then rules with the dimensionality of the pattern space.
ANN	• Ability to work with very complex distribution of objects • Works with many applications • Generality • Usefulness of the nonlinearity process of the neural network • Ability to learn by instances • Self-organizing ability • High tolerance of noisy data • Ability to work using little knowledge of the relationships between attributes • Capability of working with continuous values • Large memory storage is not required.	• Particularly slow, especially in the training phase • Difficult to determine the parameters of the network

Based on the review conducted, ANNs have been used for medical data classification problems in a wide range of research studies. The main reasons for using ANNs in a huge number of applications have been highlighted in Table 2.1.

The improved results reported in those studies that have used ANNs to perform medical data classification motivates further study to attempt to design a more enhanced ANN model in order to achieve even more accurate results in medical datamining applications.

Many different types of ANN model have been proposed in the literature. Each form of ANN has special characteristics for an exact set of conditions, similar to the practical specificity that is linked with different areas of the brain. However, all ANN models can be described in terms of three basic units: the models of the neurons, the models of the interconnections, and the training rules for renewing the weights.

The simplest and most common form of MLP is three-layered MLP consisting of an input layer, a hidden layer, and an output layer, and the number of inputs is typically predetermined to match the length of the input vector. The number of nodes in the hidden layer is selected experimentally and the number of outputs to be modelled usually decides the number of output nodes in the network. Each neuron has a number of inputs and a number of outputs. A node computes its own output based on the weighted sum of all its inputs by applying an activation function. Information flows one way in this kind of neural network from external inputs into the first layer, and they are then sent out from the first to the last of the hidden layer(s), after which they are transmitted to the output layer until the external outputs are achieved.

In recent years the ANN structure most commonly used in the literature has been the MLP Yang & Chen 2012 [15]; Zanchettin et al. [20] because it is the simplest yet most effective strategy in the area of supervised learning algorithms. The most important advantages of MLPs compared to other neural models are the approximate mapping between input and outputs and easy implementation Amiryousefi et al. [70]. A MLP consists of an input layer, an output layer and one or more hidden layer(s) that assist in capturing the nonlinearity in a system. The advantages of MLPs would be of benefit to the kind of advanced ANN that this study is attempting to develop.

2.3.2 Metaheuristic approaches

Metaheuristics consider as broadly useful algorithms that can be connected to tackle any enhancement issue.

Metaheuristic algorithms is optimized hyper search methods used to mine search space with multi-dimension when there are no specific precise certain solutions could be specifying.

Metaheuristics algorithms could be classified in several different ways: 1) According to inspiration way, (natural, biology) inspiration. 2) According to activation function implementation (static or dynamic). 3) According to the usage of memory in the process. 4) According to initial solution (single base solution or population base solution). Or else, there are different way to categories metaheuristic.

In this survey algorithms will categorize as single base solution or population base solution because this classification is better to define algorithms attributes and features, attributes and features are distinguishing more clearly. While using the other classifications may cause confuse and the algorithms features overlapping.

Metaheuristic could be divided into different categories, but the most popular division is depend on solution based Blum and Roli [13], Bossaid et al. [71].

2.3.2.1 Single based solution

In this section, we outline single-solution ion based metaheuristics, also called trajectory methods. Unlike population-based metaheuristics, they start with a single initial solution and move away from it, describing a trajectory y in the search space. Some of them can be seen as ‘‘intelligent’’ extensions of local search algorithms. Trajectory methods mainly encompass the simulated annealing method, the Tabu search, the GRASP method, the variable neighborhood search, the guided local search, the iterated local search, and their variants.

Single base solution metaheuristics are outlined in this section, likewise called trajectory methods, these techniques are direction techniques. Dissimilar to population-based metaheuristics, they begin with a solitary primary solution and move far from it, depicting a direction trajectory in the solution space. Some of them can be viewed as "smart" augmentations of neighborhood local seek algorithms. Research consider two trajectory techniques for the most part incorporate the simulated annealing method, and the Tabu search, and their variants.

1. Simulated annealing SA

Simulated annealing based on previous method called "Metropolis algorithm" Metropolis et al. [72]; Blum and Roli [13], in which some trades that do not lower the mileage are accepted when they serve to allow the solver to "explore" more of the possible space of solutions Blum and Roli [13]. Such "bad" trades are allowed using the criterion in Eq. (2.3)

$$e^{-\frac{D}{T}} > Rand(0,1) \tag{2.3}$$

where D= new solution – current solution, and T: temperature.

In following some recent related simulated annealing literature proposed in the field of medical data classification.

Ludermir [19], associates the upsides of simulated annealing, Tabu search and the backpropagation preparing methods so as to create an algorithm for delivering systems with high accuracy and more simplicity. Test comes about acquired with four characterization issues and one expectation issue has appeared to be superior to those got by the most regularly utilized advancement procedures.

Örkcü et al. [73], develop hybrid intelligent model hybridGSA (hybrid Genetic Algorithm and Simulated Annealing) for training artificial neural networks (ANN), system aims to exploit the advantages of genetic and simulated annealing algorithms and alleviate their limitations. Hybrid-GA uses three benchmark data sets Breast Cancer Wisconsin WDBC, Pima Indians Diabetes PID, and Liver Disorders LD.

Zanchettin [20], Proposed hybrid system depends on the incorporation of the simulated annealing SA, tabu search TS, genetic algorithm GA, and backpropagation, though TSa does not utilize GA. The primary points of interest of GaTSa are the accompanying: a helpful procedure to include new hubs in the engineering in light of GA, the capacity to escape from local minima with tough moves SA features, and convergence done by solution assessments of TS features.

Alwesha & Abdullah [74] developed probabilistic neural network PNN by propose a method that hybridizes the firefly algorithm with simulated annealing SA. SA control the randomness step inside the firefly algorithm while optimizing the weights of the standard PNN model. Then extend work by using Lévy flight within the firefly algorithm to explore the search space in order to improve the performance of the PNN and obtain better results in terms of classification accuracy.

2. Tabu search TS

Tabu search TS was first introduce by Glover in 1986 [75]. TS was intended to deal with local search algorithm. It expressly utilizes the historical search, to solve local minima problem and to execute an explorative technique. Its main feature is without a doubt in view of the utilization of systems brightened by the human memory. It takes a different way inverse to that of SA, which does not utilize memory, and in this manner, can't gain from the past.

Ludermir [19], associates the upsides of simulated annealing, tabu and the backpropagation preparing methods so as to create an algorithm for delivering systems with high accuracy and more simplicity. Test comes about acquired with four characterization issues and one expectation issue has appeared to be superior to those got by the most regularly utilized advancement procedures.

Zanchettin [20], Proposed hybrid system depends on the incorporation of the simulated annealing SA, tabu search TS, genetic algorithm GA, and backpropagation, though TSa does not utilize GA. The primary points of interest of GaTSa are the accompanying: a helpful procedure to include new hubs in the engineering in light of GA, the capacity to escape from local minima with tough moves SA features, and convergence done by solution assessments of TS features.

2.3.2.2 Population based solution

1. Evolutionary algorithms

Evolutionary Computation (EC) is a terminology refer to several optimization algorithms that are enlivened by the Darwinian standards of nature's ability to advance living creatures very much adjusted to their condition. Normally discovered assembled under the term of algorithms (additionally called Evolutionary Algorithms (EAs)) Boussaïd et al. [71], are the spaces of genetic algorithms, differential evolutionary, evolutionary programming, and genetic programming. Regardless of the contrasts between these systems, all offer a typical hidden thought of recreating the development of individual structures by means of procedures of selection, recombination, and mutation reproduction, in this manner delivering better arrangements [71]; [13] and [76].

- Genetic algorithm

GA searches for optimal solutions according to fitness functions by applying evolutionary algorithm to a set of populations, as in Holland [77]; Mitchell [78]; and Goldberg [79]. The algorithm functionality expressed in the selection, crossover, and mutation operations as follow:

1- Initialize primary population (normally random), in this work population represent set of NN weights with its Accuracy value as fitness value.

2- Select chromosomes to perform crossover operation.

3- Perform crossover operation according (crossover probability), and to methodology (single point or multi point crossover).

4- Perform mutation operation according to mutation probability.

5- Evaluate chromosomes by (fitness function), here (ANN) is the fitness function and Accuracy is fitness value.

6- Perform selection.

7- Repeat steps from 3 to 7 until stopping criteria.

8- Best fitness chromosome will be the solution.

Population evolutionary algorithms for searching solutions similar to genetic evolution in nature are considered in GA, section 3.2.2.1 describe GA in detail.

GA mentioned a lot in medical data mining specially as a preprocessing in feature selection, furthermore, in classification.

Mizuta et al. [80] aimed to generate the best network structure and the optimum parameter set by proposing genetic algorithms (GAs) for designing and training neural networks.

Örkcü et al. [81] utilized the advantages of genetic and simulated annealing algorithms and mitigated their limitations using a hybrid intelligent model (hybrid GA and simulated annealing) for training ANNs. Three benchmark data sets, namely, WDBC, PID, and LD, were employed to implement the model.

Salman et al. [82] study impact of metaheuristic iterations on different ANN structures. Algorithms named PSO, GA, and FW hybridized with ANN to classify PID, WDBC, LD, HSS, PD, and RNA-Seq big medical dataset, results succeed to prove the proposed methodology.

Abed et al. [83] diagnose WBCD and WDBC by hybridize KNN with GA. The proposed methodology based upon use GA as an optimization to selecting best features from the medical datasets as well as K value for KNN to obtain better accuracy.

Jabbar et al. [84] combine KNN with GA to enhance the accuracy of heart disease data set from UCI. In the context of the research GA play role for globally search and find optimal solution in a complex multimodal.

Avci and Dogantekin [85] proposed a classification method for Parkinson disease by enhance neural network with single layer using genetic algorithm, wavelet kernel, and Extreme Learning Machines. In proposed method, optimal parameter values and neurons numbers in hidden layer obtained by GA.

Shen et al. [86] tune SVM parameters by hybridize with fruit fly optimization algorithm to enhance four medical datasets classification. Proposed algorithm compared and overcome four hybridized metaheuristic algorithms with SVM on the selected medical benchmarks.

Choubey and Paul [87] proposed a methodology to facilitate PID diagnoses using GA as natural selection optimized method for feature selection in the first stage. Secondly, decision tree classify the targeted medial dataset.

Mafarja et al. [88] proposed wrapper-feature selection based on binary dragonfly algorithm as feature selection. Proposal was compare with PSO and GA. The method implemented on eighteen benchmarks including medical data sets.

Jaddi and Abdullah [89] reduced rough set attributes for thirteen datasets including some medical datasets. The approach hybridizes GA with linear and nonlinear great deluge algorithm to reduce attributes. GA used to optimize GD for attribute reduction.

Zainudden et al. [90] proposed a feature selection method based on harmony algorithm in order to improve accuracy and cost for wavelet neural network as well as with PSO and GA. Proposed method tested on ten UCI benchmarks and two real binary medical classification problems.

- Differential evolutionary algorithm

DEA initialize random population of d-dimensional vectors. The representation of the solution which applied to DEA is the same as it is applied in GA.

DEA combine several solutions with the candidate solution to generate a new solution Storn and Price [91]. Three main operations consider for population solution evolution through repeated cycle, these operations are: mutation, crossover, and selection. In spite of similarity in naming with GA operations names but they are not exactly the same. In section 4.2.2 sufficient description for topic.

Thein et al. [92], improve performance of ANN (which suffer from sluggish convergence and always being trapped at the local minima), by using differential evaluation algorithm (DEA), to determine optimal value or near optimal value for ANN parameters, and to solve issues on DE approach such as longer training time and lower classification accuracy. The proposed system implements the island-based training method to be better accuracy and less training time by using and analyzing between two different migration topologies.

Soliman, and AboElHamd [93] combine Least Squares Support Vector Machine (LS-SVM) and Differential Evolution (DE). The proposed hybrid classification algorithm which integrates DE and LS-SVM algorithms was composed of two main phases parameters Optimization and Classification phase. LS-SVM parameters were optimized by DE algorithm. method used to classify BC patients. DE guarantee the robustness of the hybrid algorithm by searching for the optimal LS-SVM parameters.

Falco [94] presents a classification tool called DEREx for medical data classification based on differential evolution and use IF-THEN rules to extract knowledge from targeted benchmarks. DERex tested on eight medical datasets and perform statistical analysis.

Abbas [95] empirically proposed an evolutionary and multi-objective classification method on breast cancer dataset the purpose of the approach is for better generalization.

2. Swarm algorithms

Swarm Intelligence (SI) is an imaginative appropriated intelligent model for taking care of optimization issues that take motivation from the aggregate conduct of a gathering of animals' social environment and activities and of other creature social orders. SI frameworks are commonly comprised of a populace of basic operators (an element equipped for performing/executing certain activities) collaborating locally with each other and with their condition. These elements with exceptionally constrained individual capacity can mutually (agreeably) perform numerous perplexing errands important for their survival. Despite the fact that there is regularly no incorporated control structure managing how singular operators ought to carry on, neighborhood collaborations between such specialists frequently prompt the rise of worldwide and self-sorted out conduct.

- Particle swarm optimization

Apart from being a stochastic, organized, and autonomous algorithm Kennedy and Eberhart [96]. PSO is a decentralized population-based evolutionary technique Ardjani et al. [97].

Basically, the algorithm consists of set of particles, this set called swarm (population of solution symbols in search space). Techniques behind the algorithm is to move particles towards the solution in the search space, this target done by choose best position gain from all particles for all the particles in the swarm.

Particles (population of solution symbols) are moved toward a solution in the search space. More details found in section 3.2.2.2. in the following some related work:

Salman et al. [82] study impact of metaheuristic iterations on different ANN structures. Algorithms named PSO, GA, and FW hybridized with ANN to classify PID, WDBC, LD, HSS, PD, and RNA-Seq big medical dataset, results succeed to prove the proposed methodology.

Shen et al. [86] tune SVM parameters by hybridize with fruit fly optimization algorithm to enhance four medical datasets classification. Proposed algorithm compared and overcome four hybridized metaheuristic algorithms with SVM on the selected medical benchmarks.

Mafarja et al. [88] proposed wrapper-feature selection based on binary dragonfly algorithm as feature selection. Proposal was compare with PSO and GA. The method implemented on eighteen benchmarks including medical data sets.

Feshki and Shijani [98] rank the effective factors of the features of the targeted medical dataset and optimized ANN by PSO for better accuracy and lest cost.

Zainudden et al. [90] proposed a feature selection method based on harmony algorithm in order to improve accuracy and cost for wavelet neural network as well as with PSO and GA. Proposed method tested on ten UCI benchmarks and two real binary medical classification problems.

Dutta et al. [99] classified five benchmark data sets from the University of California, Irvine (UCI) repository using their improved firework, LM, and PSO that included an ANN model.

- Fireworks Algorithm

FWA is established by Tan Ying in 2010 as an intelligent swarm optimization algorithm [99]. This algorithm simulates fireworks that launch spark showers that burst around them as illustrated in Tan [99] and Dutta et al. [84].

(N) fireworks denote the initial population, and the created sparks are the potential solutions around the population. Substantial sparks are produced in each iteration (firework), and a random mutation is conducted for random fireworks to maintain variants (diversification). More details found in section 3.2.2.3.

Dutta et al. [99] classified five benchmark data sets from the University of California, Irvine (UCI) repository using their improved firework, LM, and PSO that included an ANN model.

Salman et al. [82] study impact of metaheuristic iterations on different ANN structures. Algorithms named PSO, GA, and FW hybridized with ANN to classify PID, WDBC, LD, HSS, PD, and RNA-Seq big medical dataset, results succeed to prove the proposed methodology.

2.4 MAIN FINDINGS FROM LITERATURE REVIEW

In this section, the findings from the literature review are discussed to highlight the classification techniques and methodologies and performance evaluation associated with the algorithms used in medical data classification. These technologies and procedures relate to the proposed study and the solutions for enhance ANN with metaheuristic algorithms like tune parameters, balance between exploration and exploitation, relation between ANN structure and hybridize algorithms.

In term of medical data classification there are few survey studies conducted, and some of these surveys are not comprehensive. Kalantari et al. [48] this survey considers a best proposal that review state-of-the-art approaches related to medical data classification. More than seventy articles studded and reviewed from publications have highly cited. The survey outline classification methods as hybrid and single. Shrivastava et al. [100] perform a comparative study for patients of Parkinson disease in order to analyze and select a better natural inspired among other. Fatima and Pasha [101] conduct impressive machine learning survey for diseases diagnoses. The survey arranged by outline diseases and the related methods used to classify this disease.

Through the extensive review of the literature in the field of medical data classification conducted above with the surveys mention about this area we can highlight some problems and challenges beside other practical techniques used frequently:

- Based on the review conducted, ANNs have been used for medical data classification problems in a wide range of research studies. The main reasons for using ANNs in a huge number of applications have been highlighted in Table 2.1. The improved results reported in those studies that have used ANNs to perform medical data classification motivates further study to attempt to design a more enhanced ANN model in order to achieve even more accurate results in medical datamining applications.

- In recent years, MLP is most commonly ANN structure used in the literature, because it is the simplest yet most effective strategy in the area of supervised learning algorithms.

- In term of medical data classification hybrid approaches results overcome single approaches and provide accurate and efficient results in diagnoses the diseases.

- According literature the most criteria used for evaluation is the accuracy (the general true proportion of the classification), the accuracy value is the critical classification measure.

- Error rate or misclassification rate is calculate Error = 1 – Accuracy, as in Han and Kambar [4]; Suganya and Somathi [102].

- From the literature review it is clear that there are many themes in the optimization algorithms that are applied in ANN training that need to be enhanced to facilitate the ANN process as well as improve the performance of the model chosen for solving medical data classification problems. The most prevalent themes are summarized below:

  1. Solution representation to optimize the weights and structure of the ANN to find most appropriate model for solving classification and time series prediction problems;

  2. Parameter setting to achieve robust parameter designs for the optimization algorithm as settings may affect the performance of the algorithm.

  3. Balancing exploration and exploitation to develop a convergence strategy in the algorithm to provide a good balance between exploration and exploitation during the search process.

In this research, the aim and the effort is to investigate ways to improve the medical classification outcomes based on themes found in the literature that still need to be considered in current research.

3. METAHEURISTIC REPETITION INFLUENCE ON THE STRUCTURE OF THE CLASSIFICATION METHOD

3.1 INTRODUCTION

Medical data systems are analytical and diagnostic systems used to help physicians and medical centers in diseases treatment, and they are crucial for improving diagnosis and treatment. Computer researchers have been interested in this field given the vital role of medical data in human life. Physicians may refer to medical data classification, including symptoms and medical analyses on critical diseases, in making decisions. A disease data set comprises symptoms of patients as attributes along with the number of instances of such symptoms.

Health care may utilize the considerable medical data accessible. Medical centers can use data mining in their analyses to provide sufficient sources on diseases for their timely detection and prevention and to avoid the high costs incurred by medical tests [93]. Scientists have implemented various data-mining approaches to diagnose and treat diverse diseases, such as diabetes Dutta et al. [99], liver disorder Tavakkoli et al. [103], Parkinson's Shrivastava et al. 2016 [100], and cancer Örkcü et al. [81]. Artificial neural network (ANN) is extensively adopted in the classification and prediction of diseases. Mandal and Banerjee [104] used ANN with conventional backpropagation training, and ANN exhibits improved accuracy and efficiency.

Heider introduced a neural network cluster that included four subfamily networks to designate a small GTPase to a subfamily and a filter network to classify small GTPases [64]. Desell modified the structure of deep recurrent neural networks using an ant colony optimization algorithm. The weights of neural networks have been trained by ant colony optimization algorithms for continuous parameter optimization [105]. Mizuta et al. [80] aimed to generate the best network structure and the optimum parameter set by proposing genetic algorithms (GAs) for designing and training neural networks. They also introduced a fitness function that relied on output errors and simplicity in network structure. Blum and Soca solved discrete optimization problems via an ant colony optimization approach [106].

Orkcu utilized the advantages of genetic and simulated annealing algorithms and mitigated their limitations using a hybrid intelligent model (hybrid GA and simulated annealing) for training ANNs. Three benchmark data sets, namely, Wisconsin Diagnostic Breast Cancer (WDBC), Pima Indian Diabetes (PID), and Liver Disorders (LD), were employed to implement the model [81].

Seera and Limm [62], integrated a fuzzy min–max neural network into classification, regression tree, and random forest in a hybrid intelligent system. The hybrid system utilized the advantages and decreased the limitations of the fundamental models; it used the fuzzy min–max neural network to learn incrementally, the classification and regression tree to predict the network outputs, and the random forest to achieve a high classification performance.

Dutta [99] classified five benchmark data sets from the University of California, Irvine (UCI) repository using their improved firework that included an ANN model [107].

Zainudin [90] adopted a wavelet neural network as a classifier on two epileptic seizure benchmark data sets from the UCI repository, and they employed an enhanced harmony-search-based algorithm to select features.

Varma established an approach to handling boundaries of decision trees and determining split points, which employed the Gini index in diagnosis (PID) [108].

Maddouri and Eloumi [109], introduced four distinct machine-learning approaches to classify biological sequences. Luo et al. [110], discussed the recent breakthrough in big data applications in the health-care field. Guarracino [111], trained a generalized classifier with a smaller subset of points and features of original data via proposing a technique for feature selection.

The focus of most of the proposed methods and algorithms is on hybridizing ANN with one or more optimization algorithms [81] or enhancing or modifying the kernel of ANN [85]; however, they do not unify the structural change in ANN and the number of iterations of metaheuristic algorithms. The effect of the selection of different numbers of iterations for more than one structure of ANN is accordingly studied.

The remainder of this chapter is organized into five sections. Section 2 discusses the materials, methodology (including the algorithms), proposed work, and specifications of medical data sets. Section 3 presents the experimental results. Section 4 elucidates the simulation results. Section 5 concludes the new suggested methodology and provides recommendations for future work.

3.2 METHODOLOGIES

3.2.1 Backpropagation network ANN

ANN can be adopted as a classification model through mapping input data to the approximate desired output. This model includes an input layer (the layer that receives inputs), an output layer (the layer that provides outputs), and hidden layer(s) between them.

The attributes from disease data sets are input to ANN in this study. These inputs are examined in the input layer and multiplied by weights. The weights are randomly initialized relative to the neurons in the hidden layer(s), where the summation is specified in the activation function, as indicated in Eq. (3.1) and Eq. (3.2)

$$s(x) = \sum_{i=1}^{n} x_i w_i \tag{3.1}$$

The neuron output is determined after the obtained summation results from the activation function are evaluated. The following sigmoid function is adopted in the proposed model:

$$f = \frac{1}{1 + e^{-s(x)}} \tag{3.2}$$

The data sets in this study are training (40%), validation (30%), and test (30%) data sets. The trained neural network structure is used to evaluate populations, and the training set of weights (w1, …, wn) are input into the metaheuristic algorithms.

3.2.2 Metaheuristic Enhancement

This study hybridizes ANN by integrating three metaheuristic optimization algorithms, namely, GA, particle swarm optimization (PSO), and fireworks algorithm (FWA). Metaheuristic algorithms as optimization algorithms could be categorized depending on the solution into two parts; first, single based solution in which the search depicted only on one solution, second one, is a population-based solution in which search procedure performed on a population of solutions. Above algorithms are categorized as population-based metaheuristic algorithms.

3.2.2.1 Genetic algorithms GA

GA searches for optimal solutions according to fitness functions by applying evolutionary algorithm to a set of populations, as in Holland [77], Mitchell [78], and Goldberg [79]. Figure 3.1 shown the algorithm functionality which expressed in the selection, crossover, and mutation operations as follow:

1- Initialize primary population (normally random), in this work population represent set of NN weights with its Accuracy value as fitness value.

2- Select chromosomes to perform crossover operation.

3- Perform crossover operation according (crossover probability), and to methodology (single point or multi point crossover).

4- Perform mutation operation according to mutation probability.

5- Evaluate chromosomes by (fitness function), here (ANN) is the fitness function and Accuracy is fitness value.

6- Perform selection.

7- Repeat steps from 3 to 7 until stopping criteria.

8- Best fitness chromosome will be the solution.

Population evolutionary algorithms for searching solutions similar to genetic evolution in nature are considered in GA.

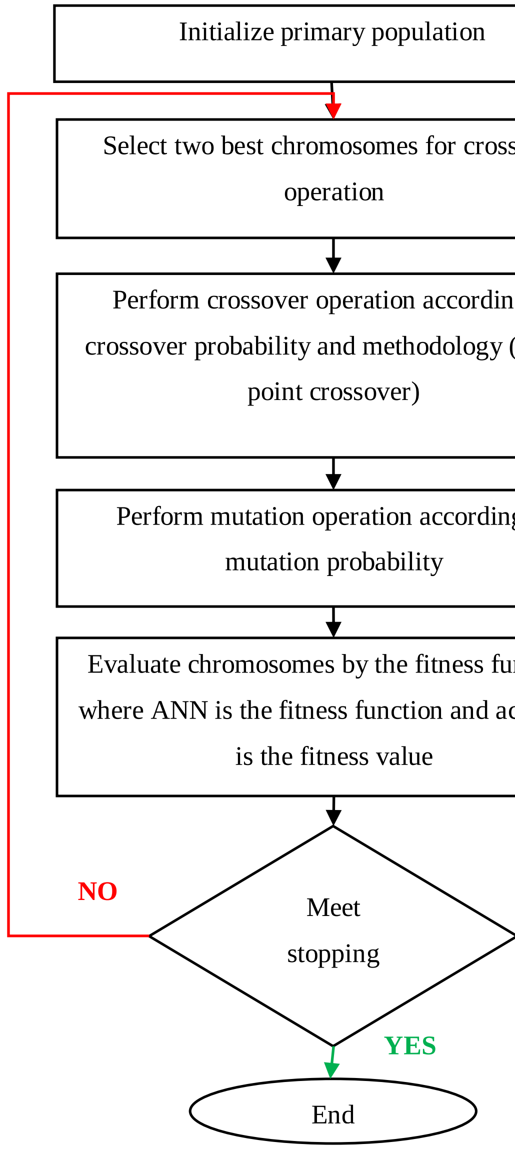


Figure 3.1: Genetic algorithm (GA)

3.2.2.2 Particle swarm optimization PSO

Apart from being a stochastic, organized, and autonomous algorithm Kennedy and Eberhart [96]. PSO is a decentralized population-based evolutionary technique Ardjani et al. [97].

Basically, the algorithm consists of set of particles, this set called swarm (population of solution symbols in search space). Techniques behind the algorithm is to move particles towards the solution in the search space, this target done by choose best position gain from all particles for all the particles in the swarm.

Particles (population of solution symbols) are moved toward a solution in the search space.

Velocity vector defines the position change of a set of particle swarms (n) in each iteration process (i). The movement direction and distance of each particle are governed by this vector, as indicated below in Eq. (3.3) and Eq. (3.4):

$$\begin{aligned}\mathrm{velocity}(\mathrm{i}+1) \\ &= \omega * \mathrm{velocity}(\mathrm{i}) + \varphi 1 \\ &* \left(\mathrm{bestLocalPostion} - \mathrm{postion}(\mathrm{i})\right) \quad + \varphi 2 \\ &* \left(\mathrm{bestGlobalPosition} - \mathrm{position}(\mathrm{i})\right)\end{aligned} \tag{3.3}$$

$$position(i+1) = position(i) + velocity(i+1) \tag{3.4}$$

where i is the number of iterations, ω is the velocity coefficient or scale that is normally 1 and decreases during iterations, φ1 is the fixed scale of difference between current and local positions, and φ2 is the fixed scale of difference between current and global positions.

Figure 3.2 illustrates the PSO process.

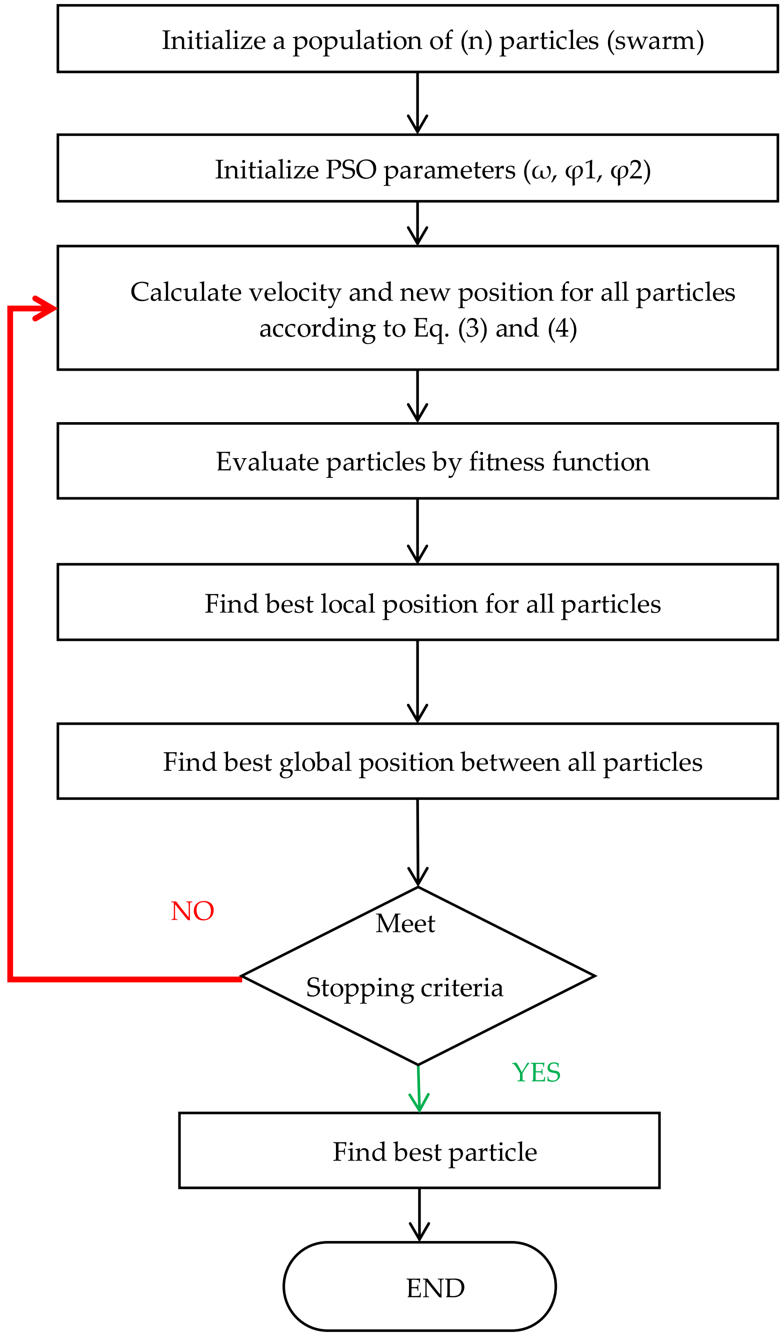


Figure 3.2: Particle swarm optimization (PSO)

3.2.2.3 Fireworks algorithm FWA

FWA is established by Tan Ying in 2010 as an intelligent swarm optimization algorithm [112]. This algorithm simulates fireworks that launch spark showers that burst around them as illustrated in Tan [112] and Dutta et al. [99].

(N) fireworks denote the initial population, and the created sparks are the potential solutions around the population. Substantial sparks are produced in each iteration (firework), and a random mutation is conducted for random fireworks to maintain variants (diversification).

The fireworks then generate sparks with different magnitudes and directions.

The maximum number of sparks (M) created from the fireworks is determined by the following Eq. **(3.5)**

$$S_i = M * \frac{f(X_i)}{\sum_{i=1}^{N} f(X_i)} \tag{3.5}$$

where Si is the number of sparks for firework (Xi), M is the maximum number of sparks, and f(Xi) is the activation function of firework (Xi).

The explosion spark amplitude can be determined in Eq. **(3.6)**

$$\mathrm{A_i} = \mathrm{A} * \frac{\mathrm{F}(\mathrm{X}_{(\mathrm{N-i+1})})}{\sum_{\mathrm{i=1}}^{\mathrm{N}} \mathrm{F}(\mathrm{X_i})} \tag{3.6}$$

where Ai is the explosion spark amplitude for firework (Xi), A is the maximum amplitude, and F(Xi) is the activation function of firework (Xi).

The FWA steps are illustrated in Figure 3.3.

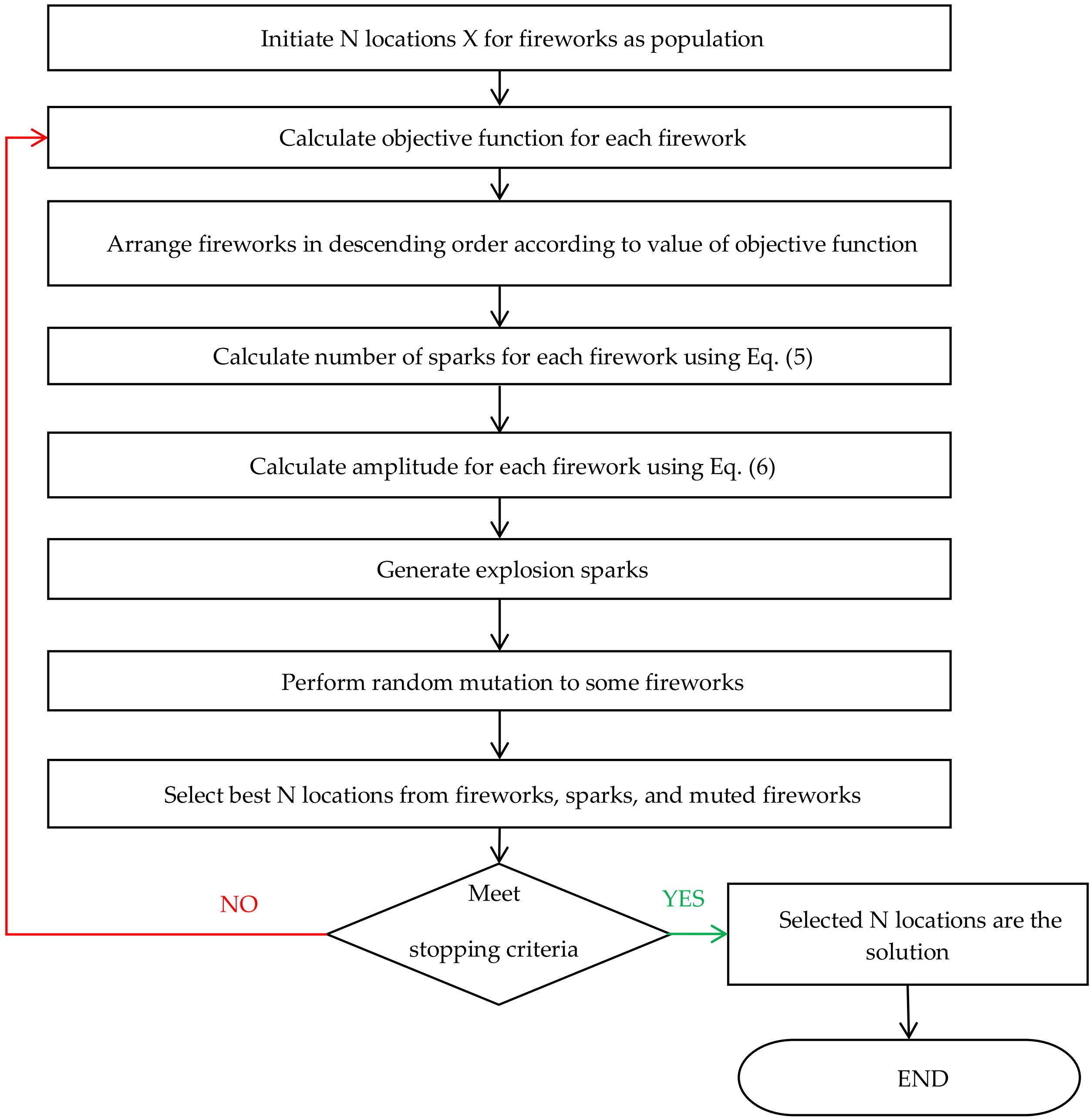


Figure 3.3: Fireworks algorithm (FWA)

3.2.3 Suggested Methodology

This study hybridizes ANN with three metaheuristic methods, namely, GA, PSO, and FWA. The best set of weights for ANN is determined for constructing these methods to produce an improved accuracy performance. The general structure of this study is presented in Figure 3.4.

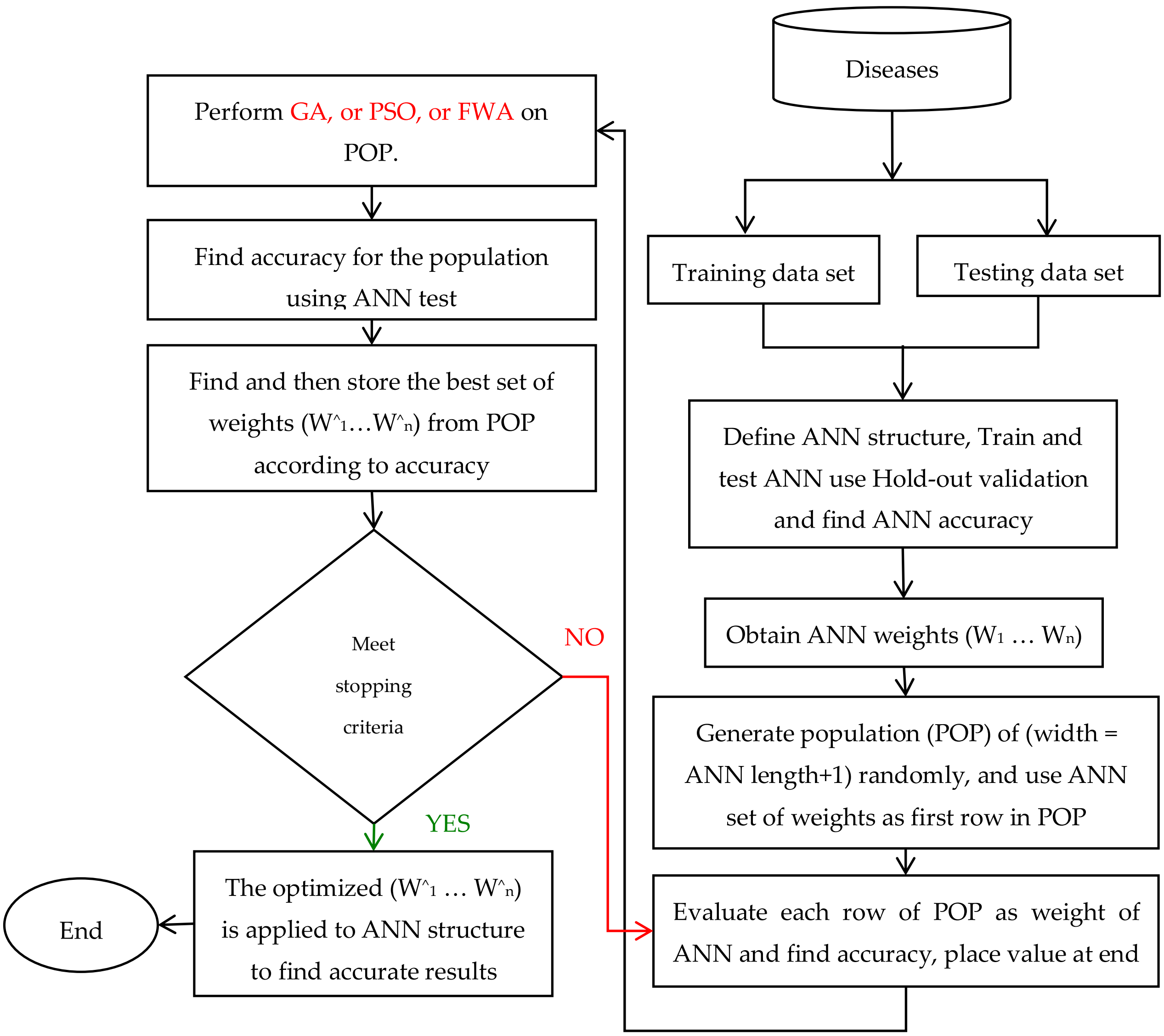


Figure 3.4: Flowchart of optimized classification by an Artificial Neural Network (ANN) with metaheuristic algorithms

ANN is trained and tested in the first phase, and the set of weights (w1, …, wn) is then obtained. Optimization algorithms are used in the second phase to obtain an improved set of weights (w^1,

…, w^n) for the trained ANN. This new set of weights (w^1, …, w^n) should achieve high accuracy with the proposed ANN structure.

3.2.4 Diseases Data Sets

Classification methods and optimization algorithms are hybridized on the following disease data sets [107].

(1) PID

This data set comprises patients who are pregnant females at least 21 years old and of the heritage from Indian Pima, according to the National Institute of the Diabetes and the Digestive and Kidney Diseases.

Number of tuples: 768

Number of attributes: 9 (including class)

(2) WDBC

The characteristics of the cell nuclei that exist in the image are described in this data set. The University of Wisconsin employs a digitized image of a fine needle aspirate of a breast mass to compute features.

Number of instances: 569

Number of attributes: 32

(3) LD

Seven attributes are included in this data set. Blood test results related to liver disorders due to alcohol consumption are indicated in the five attributes. The number of drinks per day constitutes the sixth attribute. Patient class and condition (i.e., whether the patient has the disorder or not) are presented in the seventh attribute.

Number of instances: 345

Number of attributes: 7 (including class attribute)

(4) Haberman Surgery Survival (HSS)

The Billings Hospital of the University of Chicago conducted a study on the survival of patients who had undergone surgery for breast cancer between 1958 and 1970. Results were gathered in this data set.

Number of instances: 306

Number of attributes: 4 (including class attribute)

(5) Parkinson's (PD)

Max Little of the University of Oxford created this data set by collaborating with the National Center for Voice and Speech in Denver, Colorado. Twenty-three attributes, which are sound measures, are contained in this data set. A total of 197 instances that correspond to sound records of 31 individuals, 23 of whom have Parkinson's disease, are also provided.

Table 3.1 provides the specifications of the data sets.

Table 3.1: Specifications of disease datasets and their attributes' and classes' numbers

DATA SET	NUMBER OF INSTANCES	NUMBER OF ATTRIBUTES	CLASS 1 (0) value	CLASS 2 (1) value
Pima Indian Diabetes (PID)	768	8	(500) not infected	(268) infected
Wisconsin Breast Cancer (WDBC)	569	31	(357) benign	(212) malignant
Liver Disorder (LD)	345	7	(145) not	(200) disorder
Haberman Surgery Survival (HSS)	306	3	(225) lived 5 years	(81) deceased
Parkinson's (PD)	195	23	(48) normal	(147) abnormal

3.2.5 Solving Strategy

The ANN architecture is tuned with specific optimization algorithm parameters and number of iterations in this study.

Improved ANN performance and enhanced accuracy in disease diagnosis problems through change in the neural network structure by tuning metaheuristic algorithm parameters and increase in the number of iterations are aimed to prove in this study.

This study mainly comprises two experiments. First, the best among GA, PSO, and FWA is determined. Second, the selected algorithm is used, in which the specified algorithm parameters are tuned to acquire enhanced results. The mean differences between ANN and (ANN + PSO) in the five benchmarks are determined via a statistical t-test. The statistical significance of the method is identified using the conditional probability p value generated from the t-test. A big data set with multiple classes is adopted to test the hybridized algorithms for determining the algorithms' performances.

Several tests are performed, and the algorithm parameters are set as follows:

(1) GA

Population = 10

Probability of crossover = 0.7

Probability of mutation = 0.1

(2) PSO

Swarm size = 100

Velocity scalar coefficient (w) = 1.0

Velocity change in each iteration = 0.99

First velocity equation coefficient (c1) = 2.0

Second velocity equation coefficient (c2) = 2.0

(3) FWA

Number of fireworks = 50

Number of sparks = 5

Maximum function evaluation = 50,000

Gaussian number for mutation = 5

Probability of mutation = 0.3

Spark upper bound = 10

Spark lower bound = −10

3.3 RESULTS

3.3.1 Hybridization Experiment

Change in ANN structure (number of hidden layers) is tested in this experiment. Five hundred and 1000 iterations are conducted for two and three of the hidden layers, respectively, using the three proposed hybridization algorithms on the five disease benchmark data sets.

The experiment design is as follows:

- Hybridizing ANN with GA, PSO, and FWA for five data sets (two hidden layers for ANN with 500 iterations) (Table 2)
- Hybridizing ANN with GA, PSO, and FWA for five data sets (two hidden layers for ANN with 1000 iterations) (Table 3)
- Registering the improvements between using 500 and 1000 iterations of GA, PSO, and FWA in two hidden layers for the five data sets
- Hybridizing ANN with GA, PSO, and FWA for five data sets (three hidden layers for ANN with 500 iterations) (Table 4)

- Hybridizing ANN with GA, PSO, and FWA for five data sets (three hidden layers for ANN with 1000 iterations) (Table 5)
- Registering the improvements between using 500 and 1000 iterations of GA, PSO, and FWA in three hidden layers for the five data sets

3.3.2 Results of the Hybridization Experiment

Table 3.2 and Table 3.3 present the results of implementing the three proposed algorithms on the benchmarks.

Table 2. Two-hidden-layer ANN hybridized with optimization algorithms GA, PSO, and FWA for 500 iterations on UCI datasets: Pima Indian Diabetes PID, Wisconsin Breast Cancer WDBC, Liver Disorder LD, Haberman Surgery Survival HSS, and Parkinson's PD.

Table 3.2: Two-hidden-layer ANN hybridized with optimization algorithms GA, PSO, and FWA for 500 iterations on UCI datasets: Pima Indian Diabetes PID, Wisconsin Breast Cancer WDBC, Liver Disorder LD, Haberman Surgery Survival HSS, and Parkinson's PD

2 Hidden_ 500 Iteration		PID	WDBC	LD	HSS	PD
Classical ANN		74.35	97.17	57.10	72.88	75.38
Hybridized	ANN + GA	78.78	97.72	72.75	78.10	90.26
	ANN + PSO	**79.82**	97.89	**74.52**	**78.76**	**90.77**
	ANN + FWA	79.56	**98.07**	73.07	78.76	89.23

Table 3.3: 2-hidden–layer ANN hybridized with optimization algorithms GA, PSO, and FWA for 1000 iterations

2 Hidden_1000 Iteration		**PID**	**WDBC**	**LD**	**HSS**	**PD**
Classical ANN		74.35	97.17	57.10	72.88	75.38
Hybridized	ANN + GA	79.40	98.07	74.52	78.10	90.77
	ANN + PSO	79.86	**98.42**	**74.52**	**79.08**	**91.28**
	ANN + FWA	**79.95**	98.07	73.33	78.43	91.28

Table 3.2 and Table 3.3 indicate the following improvements:

- ANN + GA: Four out of five benchmarks
- ANN + PSO: Four out of five benchmarks
- ANN + FWA: Three out of five benchmarks

Table 3.4: 3-hidden–layer ANN hybridized with optimization algorithms GA, PSO, and FWA for 500 iterations

3 Hidden_ 500 Iteration		**PID**	**WDBC**	**LD**	**HSS**	**PD**
Classical ANN		76.48	95.39	56.81	73.57	80.51
Hybridized	ANN + GA	78.39	97.89	71.59	77.78	90.77
	ANN + PSO	**79.82**	98.24	**73.91**	**78.43**	**89.74**
	ANN + FWA	78.39	**98.42**	70.14	77.78	89.28

Table 3.5: 3-hidden–layer ANN hybridized with optimization algorithms GA, PSO, and FWA for 1000 iterations

3 Hidden_ 1000 Iteration		**PID**	**WDBC**	**LD**	**HSS**	**PD**
Classical ANN		76.48	95.39	56.81	73.57	80.51
Hybridized	ANN + GA	78.91	98.07	71.59	78.10	90.77
	ANN + PSO	**80.21**	98.07	**75.65**	**78.76**	**91.28**
	ANN + FWA	79.82	**98.42**	73.91	78.10	90.77

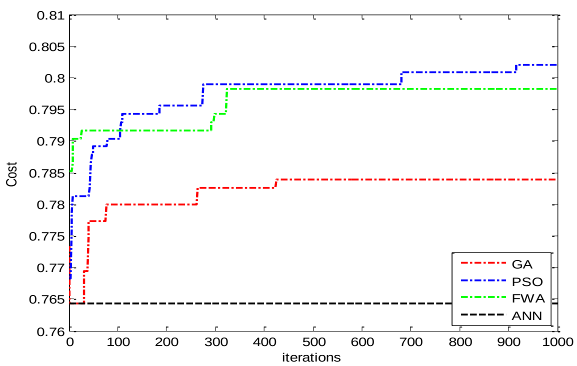
Table 3.4 and

3 Hidden_ 500 Iteration		**PID**	**WDBC**	**LD**	**HSS**	**PD**
Classical ANN		76.48	95.39	56.81	73.57	80.51
Hybridized	ANN + GA	78.39	97.89	71.59	77.78	90.77
	ANN + PSO	**79.82**	98.24	**73.91**	**78.43**	**89.74**
	ANN + FWA	78.39	**98.42**	70.14	77.78	89.28

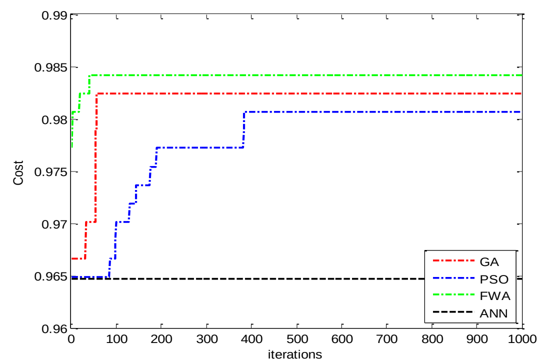
Table 3.5 demonstrate the following improvements:

- ANN + GA: Three out of five benchmarks are improved.
- ANN + PSO: Four out of five benchmarks are improved.
- ANN + FWA: Four out of five benchmarks are improved.

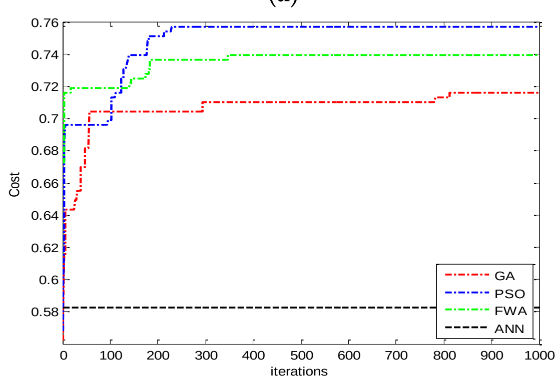
The performance of each proposed algorithm in enhancing ANN with three hidden layers using 1000 iterations is illustrated in Figure 3.5.The best performance in most cases is given by the PSO with three-hidden-layer ANN and 1000 iterations. Therefore, the most promising algorithm relative to enhancement is the hybridized PSO with ANN.



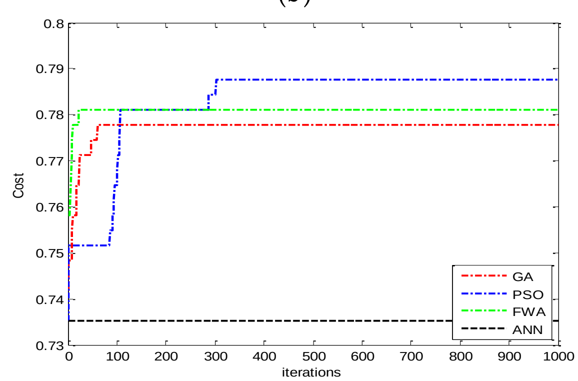
(**a**)



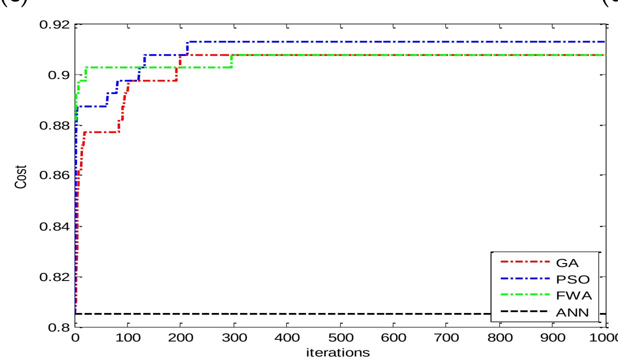
(**b**)



(**c**)



(**d**)



(**e**)

Figure 3.5: Shows the experiment, the classification performance of the three proposed hybrid algorithms GA, PSO, and FWA applied on the five diseases data sets: (a) PID; (b) WDBC; (c) LD; (d) HSS; (e) PD University

The means of two algorithms are compared by a t-test, which is a statistical hypothesis test. Correlation and regression are applied to determine the differences of the two algorithms. The t-test statistics is converted into a conditional probability called p-value. The question, "If the null hypothesis is true, then what is the probability of observing the current data or data that are more extreme?" is answered by the p-value.

The observed data are unlikely when the null hypothesis is true; therefore, a small p-value provides evidence against the null hypothesis Afifi and Azen [115].

Significance is tested by setting a default value (0.05) as an indicator, and ($p < 0.05$) is considered significant.

The following formula Eq. (3.7) is integral to the t-distribution probability density function and is used to compute p-value:

$$\frac{1}{\sqrt{\mathrm{df}\beta(\frac{1}{2},\frac{\mathrm{df}}{2})}}\int_{-t}^{t}\left(-\frac{x^2}{\mathrm{df}}\right)-\frac{\mathrm{df}+1}{2}\,dx \tag{3.7}$$

where df pertains to the degree of freedom and β denotes the beta function.

The statistical test on the five UCI benchmarks is significant on the basis of determining the differences between ANN and optimized ANN + PSO, as shown in Table 3.6.

Table 3.6: Statistical test and of 5 University of California in Irvine (UCI) datasets: Pima Indian Diabetes PID, Wisconsin Breast Cancer WDBC, Liver Disorder LD, Haberman Surgery Survival HSS, and Parkinson's PD, between ANN and ANN + PSO

	PIMA	**WDBC**	**LD**	**HSS**	**PD**
Mean difference	0.02889000	0.00929500	0.17477500	0.02254000	0.03258000
T score	16.3009	9.9128	64.9535	30.3677	5.3323
standard error of difference	0.002	0.001	0.003	0.001	0.006
Difference verdict	Significant	Significant	Significant	Significant	Significant

3.4 DISCUSSION

An effective factor in ANN learning is the number of hidden layers. An increase in the number of these layers leads to increased ANN complexity and overfitting in ANN learning. Nevertheless, when this number is increased, the model can obtain precise classification results. Metaheuristic algorithms are employed for ANN generalization to solve the overfitting problem and enhance the classification.

Metaheuristic algorithms, which are search algorithms, optimize mathematical models to yield better solutions Talbi [76] and Blum [13]. Different strategies are iteratively adopted in these algorithms. For instance, algorithm divergence can be prevented by accurately tuning parameters within a single iteration of the algorithm, and mathematical models (the ANN model in this study) can search for better solutions within the search space by increasing the number of iterations.

The following observations are acquired from hybridization experiment:

- Tables 2 and 3 indicate that the following effects can be achieved by increasing the number of iterations in two hidden layers from 500 to 1000:

(1) ANN + GA: Four out of five benchmarks are improved.

(2) ANN + PSO: Four out of five benchmarks are improved.

(3) ANN + FWA: Three out of five benchmarks are improved.

- Tables 2 and 3 imply that the following effects can be produced by increasing the number of iterations in three hidden layers from 500 to 1000:

(1) ANN + GA: Four out of five benchmarks are improved.

(2) ANN + PSO: Four out of five benchmarks are improved.

(3) ANN + FWA: Three out of five benchmarks are improved.

The hybrid ANNPSO algorithm is determined to be the most suitable methodology for this study according to the hybridization experiment.

The best algorithm from the hybridization experiment is used and the parameters of the PSO algorithm are fine tuned in this experiment, the population initialization for the three algorithms is divided into two processes. First, initialization is conducted with the ANN set of weights (w_1…, w_n) and around neighbors. Second, random initialization is performed.

Better solutions in the hyperspace can be gain via population density in the future.

PSO is determined to be the best algorithm in this study according to the aforementioned results. The results of this algorithm are comparable to those of some state-of-the-art algorithms, as indicated in Table 3.7. Figure 3.6 depicts the enhancement in the PSO algorithm with ANN on the five medical data sets.

Table 3.7: Comparative results

Pima Indian Diabetes (PID)	77.60% (Au et al., 2001)	[33]
	75.29% (P. Luuka, 2009)	[34]
	77.60% (Örkcü, 2015)	[5]
	80.21% (proposed)	
Wisconsin Breast Cancer (WDBC)	94.00% (Örkcü, 2011)	[35]
	97.49% (P. Luuka, 2011)	[36]
	97.29% (Seara & Lim, 2014)	[11]
	98.42% (proposed)	
Liver Disorder (LD)	70.25% (Luuka, 2009)	[34]
	74.86% (Lee and Mangasarian, 2001)	[37]
	75.65% (proposed)	
Haberman Surgery Survival (HSS)	72.70% (Pham, 2011)	[38]
	51.96% (Yang et al., 2017)	[39]
	78.76% (proposed).	
Parkinson's (PD)	85.03% (p. Luuka, 2011)	[36]
	81.34% (Rustempasic & Can 2013)	[40]
	93.60% (Shrivastava 2017)	[4]
	92.28% (proposed).	

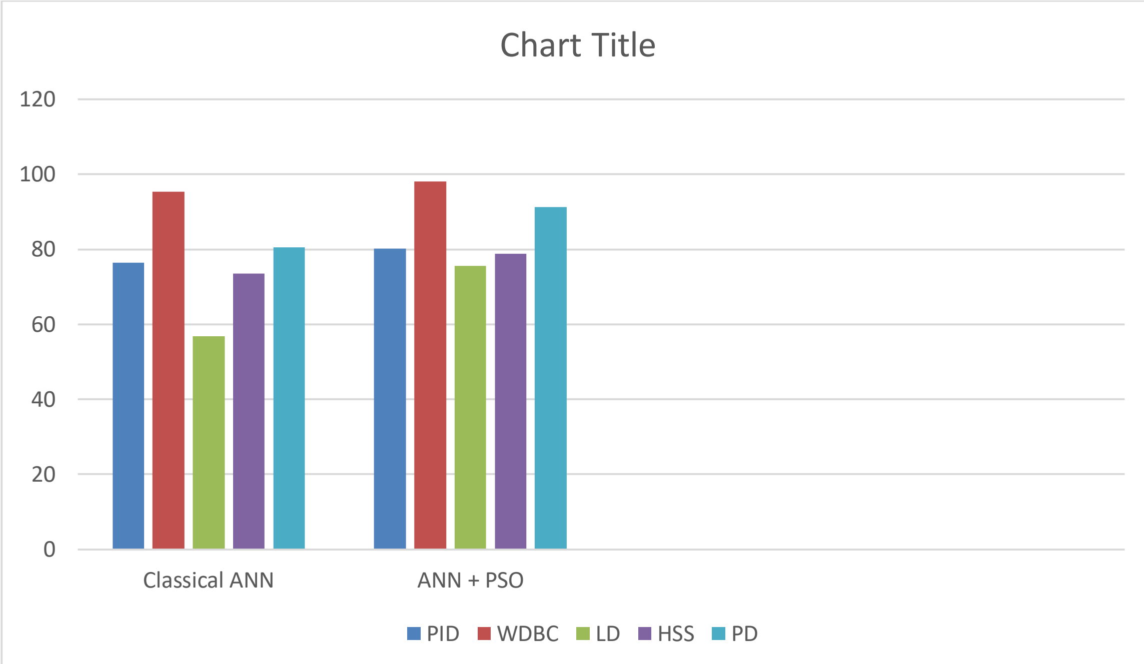


Figure 3.6: Bar charts show accuracy enhancement of hybridization experiment.

Significant differences between the accuracies of ANN and ANN + PSO on the five UCI data sets are determined in the statistical test.

The aforementioned results do not necessarily indicate that the hybridized PSOANN is a universal approach, and this notion is incorrect. The no–free-lunch theorem NFLT (Wolpert and Macready [35]; Ho and Pepyne [36]) states that no superior optimization algorithm exists for all problems. This finding is could be acquired from implementing the same optimization algorithms on the another medical repository in the future to prove NFLT.

The effects of population-metaheuristic number of iterations on varying ANN structures and tune-metaheuristic parameters are evaluated in this study. The capability of the population algorithms to classify the large biomedical data set is tested. The findings of this study can be extended to different fields and applications in future work. For instance, the effect of selecting different ANN kernel functions on hybridization or on applying various types of metaheuristic algorithms, such as trajectory algorithms, can be studied. Metaheuristic and machine learning can be hybridized to improve metaheuristic mathematical models for balancing exploration and explanation.

3.5 CONCLUSIONS

ANN is hybridized with three optimization algorithms in this study. They are GA, PSO, and FWA. PSO is adopted to hybridize ANN with diverse layer structures, and its parameters are tuned. Unlike most of other approaches in the literature, PSO is competitive on the five utilized medical data sets Table 3.7. Our approach can be adopted in future studies to solve various medical problems.

4. IMPROVING ANEMIA CLASSIFICATION BY EVALUTIONARY TRAJECTORY METAHEURISTIC HYBRIDIZATION

4.1 INTRODUCTION

In this chapter ANN hybridized with hybridization between differential evaluation algorithm (DEA) and simulating annealing (SA) to improve anemia disease classification. Both DEA and SA are metaheuristic algorithms, DEA consider evolutionary population-based solution algorithm while SA is a trajectory method work on single solution. Balance between diversification and intensification is the major role of metaheuristic algorithms Yang et al. [116], Blum and Roli [13], and Talbi [76]. Diversification (exploration) performed well with evolutionary metaheuristic algorithms, while trajectory algorithms consider best in local search (explanation). In this work hybridization between Differential evolution (DEA) and simulating annealing (SA) to enhance artificial neural network for Anemia medical data set classification. Proposed methodology begins with choosing best ANN structure (best number of hidden layers and best number of neurons in each layer) then classification process enhanced by hybridizing DEA and SA to obtain batter classification accuracy. Proposed methodology registered significant enhancement in Anemia classification, Anemia data is real dataset gathered from Iraqi blood laboratories to detect Anemia diseases. In addition, the proposal applies into two bench marks from university of California (UCI) repository, pima Indians diabetes PID and liver disorder LD diseases. Proposal method registers remarkable results.

There are many different data mining algorithms in literature used to classify several types of diseases, such as anemia disease for specific types based on Data Mining algorithms Elshami & Alhalees [117]. A person with anemia probably unaware of the problem, because symptoms may not appear. Millions of people may have anemia and their health exposed risk. Therefore the disease is significant, several studies carried out in this domain mentioned in the literature Yilmaz et al. [118]. Sanap et al. [119] developed a system using the classification technique C4.5 decision tree algorithm and SMO support vector machine WEKA. They implemented a number of experiments using these algorithms. The anemia classification using decision tree that given clear results depend on CBC reports. Amin et al. [120], have compared between naive Bayes, J48 classifier and neural network classification algorithms using WEKA and working on hematological data to specify the best appropriate algorithm.

There is no rule or method guarantee balance between exploration (diversification) and explanation (intensification) in metaheuristic algorithms for all problems Yang et al. [116]. However, there are different studies try to solve balancing between diversification and intensification. Fagan and Vuuren [121] declare six general views of diversification and intensification terminology from literature reviews. Another study Makas and Yumusak [122] combine artificial bee colony ABC with migrating birds optimization MBO in order to perform balance between exploration and explanation via utilize exploration property of ABC and explanation property of MBO by means of a sequential execution strategy.

4.2 MATERIALS AND METHODS

4.2.1 Artificial neural network ANN

ANN can be adopted as a classification model through mapping input data to the approximate desired output. This model includes an input layer (the layer that receives inputs), an output layer (the layer that provides outputs), and hidden layer(s) between them.

The attributes from disease data sets are input to ANN in this study. These inputs are examined in the input layer and multiplied by weights. The weights are randomly initialized relative to the neurons in the hidden layer(s), where the summation is specified in the activation function, as indicated in Eq. (3.1) and Eq. (3.2)

$$s(x) = \sum_{i=1}^{n} x_i w_i \tag{4.1}$$

The neuron output is determined after the obtained summation results from the activation function are evaluated. The following sigmoid function is adopted in the proposed model:

$$f = \frac{1}{1 + e^{-s(x)}} \tag{4.2}$$

The data sets in this study are training (40%), validation (30%), and test (30%) data sets. The trained neural network structure is used to evaluate populations, and the training set of weights (w1, …, wn) are input into the metaheuristic algorithms.

4.2.2 Differential Evolution Algorithm DEA

DEA initialize random population of d-dimensional vectors. The representation of the solution which applied to DEA is the same as it is applied in GA.

DEA combine several solutions with the candidate solution to generate a new solution Storn and Price [91]. Three main operations consider for population solution evolution through repeated cycle, these operations are: mutation, crossover, and selection. In spite of similarity in naming with GA operations names but they are not exactly the same.

The process of generation in every iteration is going to be like that Storn and Price [91]; Kachitvichyanukul [123]:

Three vectors are selected randomly from the population (not the target vector) and combine to generate mutant vector V as first step. combine process performed by the following Eq. (4.3).

$$V = X_1 + f(X_2 - X_3) \tag{4.3}$$

Where X_1, X_2, and X_3 are the random selected vectors from the population and F is a constant factor controls the amplification of the differential variation (X_2 – X_3) and consider as main parameter of DEA.

Second step is crossover between mutant vector and the target vector. Crossover method in DEA used either binomial crossover Eq. (4.4) to produce trial vector according specified crossover probability CR.

$$U_i = \begin{cases} V_i, & random(i) \leq CR \\ X_i, & random(i) > CR \end{cases} \tag{4.4}$$

Where U is trial vector, V is mutant vector, and X is the target vector. CR is the crossover rate, while i: 1 ... D, D consider the number of dimensionality. Figure 4.1 show crossover process for 8-dimension parameters.

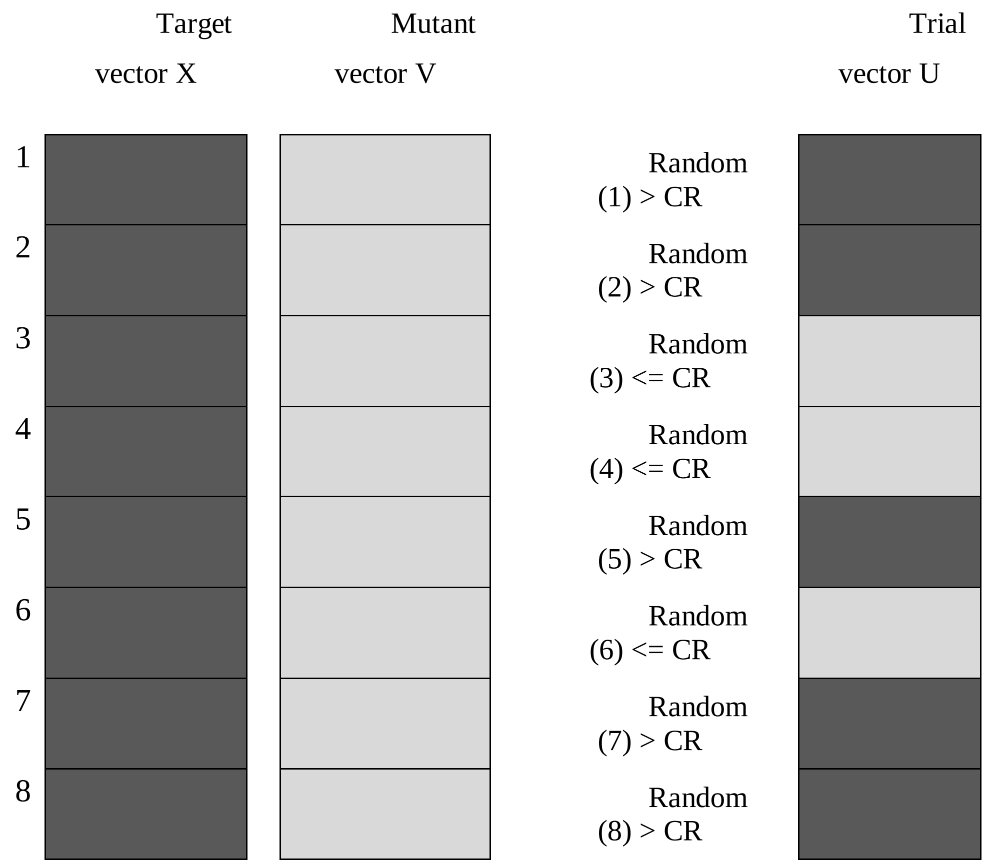


Figure 4.1: Crossover process for 8 parameters

Third step is selection operation that is choose best vector between target vector and trial vector according fitness. Best fitness will be target in the next generation.

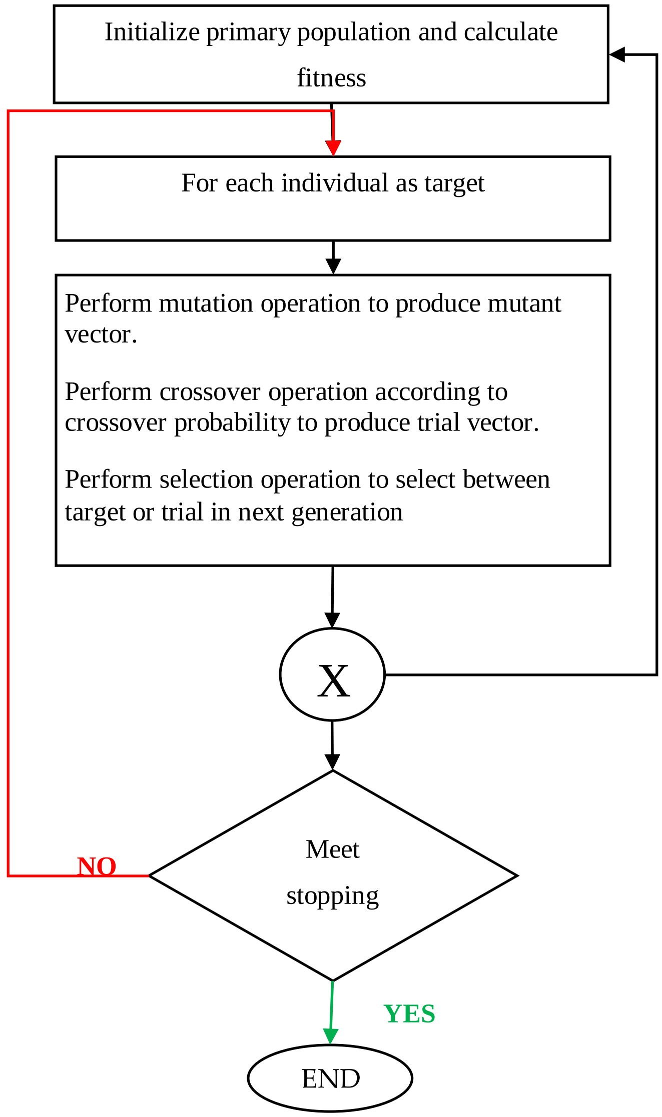


Figure 4.2 illustrate operation steps of differential evolution algorithm as flowchart.

In term of this study each individual will represent the set of weights for ANN model and fitness would be ANN accuracy.

Figure 4.2: Differential evolution flowchart DE

4.2.3 Simulated annealing SA

Simulated annealing uses a single agent or solution which moves through the design space or search space in a piecewise style Kirkpatrick et al. [124]. Basically, simulated annealing based on previous method called "Metropolis algorithm" Metropolis et al. [72]; Blum and Roli [13], in which some trades that do not lower the mileage are accepted when they serve to allow the solver to "explore" more of the possible space of solutions Blum and Roli [13]. Such "bad" trades are allowed using the criterion in Eq.(4.5):

$$e^{-\frac{D}{T}} > Rand(0,1) \tag{4.5}$$

where D= new solution – current solution, and T: temperature.

General form for trajectory methods could be formulated as follows:

```
---------------------------------------------------
Initialize solution
Loop
     Improve
Until stop criteria
-------------------------------------------------
```

In term of this study, simulated annealing algorithm steps would be like the following:

```
Sol.w ← initialize solution
Sol.Acc ← Eval(sol.w)
Sbest ← Sol
Initialize  ( Iter    Temp   Alpha) values
For i= 1:Iter
Newsol.w  ← create neighbor solution  (Sol.w)
newsol.Acc=Eval(newsol.w)
if ( newsol.Acc > sol.Acc)
sol=newsol
else if   (exp((sol-newsol)/Temp) > Rand)  % (sol-newsol) is a negative value
            sol=newsol
     end %if
end %if
```

if sol.Acc > Sbest.Acc

Sbest=sol

Temp=Alpha*Temp % Alpha is constant used to decrease Temp

End % FOR

The search will depend on the value of $\boldsymbol{e}^{(-\frac{newsol-sol}{Temp})}$ value, which decrease over time to provide diversification for the solution through algorithm loop iteration.

4.2.4 Proposed Work

This study presents the details of hybridization between Differential evolution (DE) and simulating annealing (SA) to enhance artificial neural network and proposed ANN_DESA method in order to balance between exploration (diversification) and explanation (intensification). Proposed methodology begins with choosing best ANN structure best number of hidden layers and best number of neurons in each layer via test every ANN structure on the dataset with boundaries of 30 hidden layer and 10 neurons to find best structure which have the best accuracy. Weights of the best structure (W1 … Wn) introduce to hybridized DESA, in which DE turn to exploration role in the method in term of diversification solutions, while SA maintain diversification role for exploitation in promising solution search space. Resulting weights (W^1 … W^n) would be tested on the choosing dataset for better accuracy. Proposal study illustrated in Figure 4.3.

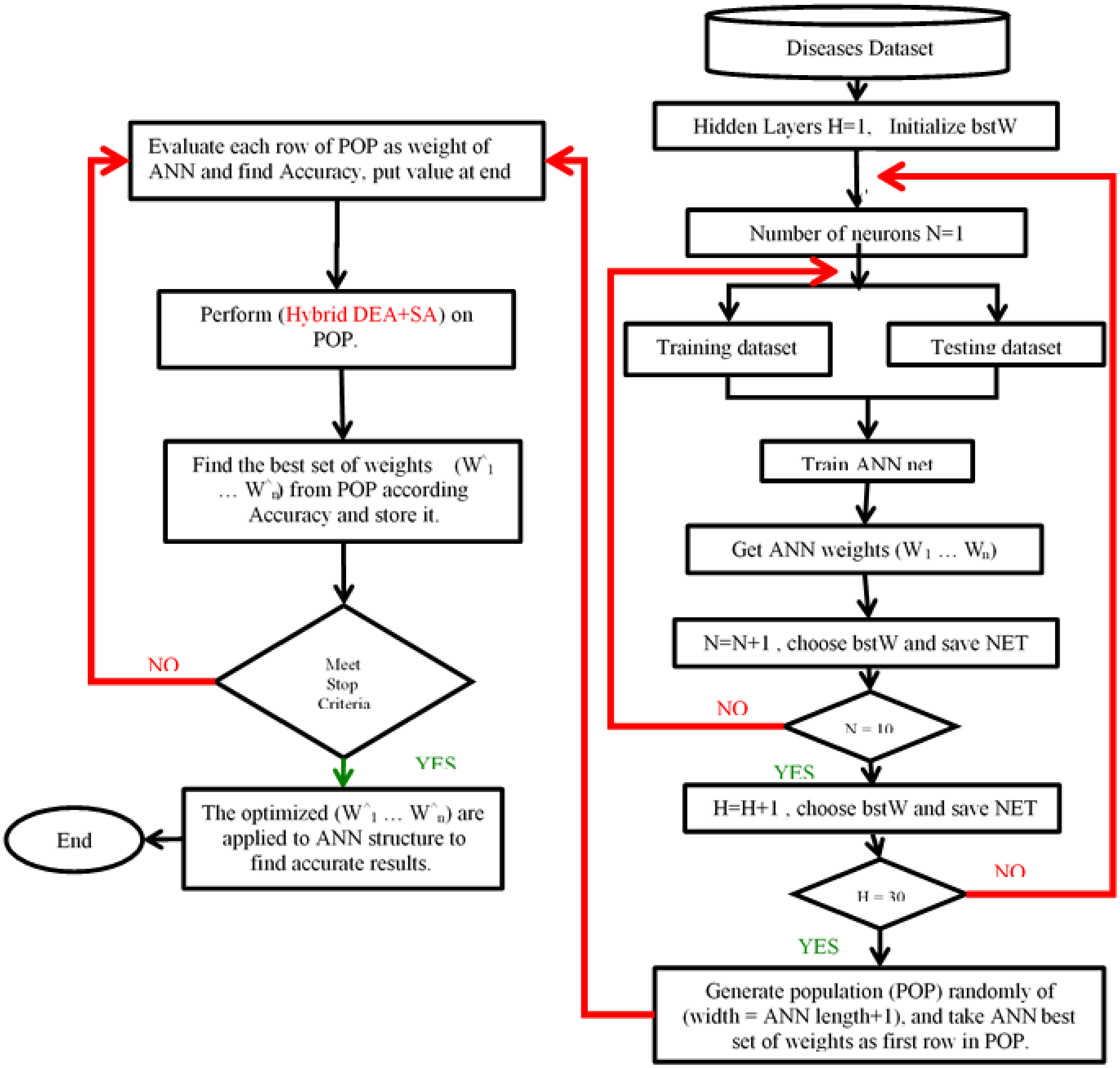


Figure 4.3: Proposal methodology

Parameter settings are illustrating in

Table 4.1, the setting for the proposed algorithms TS, SA, GA, DE, and DESA when tested on Anemia, PID, and LD.

Table 4.1: Parameters setting for proposed work

Differential Evolution DE		**Genetic Algorithm**	
Parameters	Value	Parameters	Value
Number of generations	1000	Number of iterations	1000
Population size	50	Population size	50
Crossover rate	0.8	Crossover rate	0.7
		Mutation rate	0.3
Tabu Search TS		**Simulated Annealing**	
Parameters	Value	Parameters	Value
Number of iterations	1000	Number of iterations	1000
		Temp	0.025
		Alpha	0.99
DESA Method setting			
DEA part parameters setting		SA part parameters setting	
Parameters	Value	Parameters	Value
Number of generations	1000	Number of iterations	100
Population size	50	Temp	0.025
Crossover rate	0.8	Alpha	0.99

In order to effectively evaluate proposal work ANN_DESA we test the three datasets with ANN hybridized with two trajectory methods SA and tabu search TS, and with two evolutionary algorithms GA and DE, finally ANN optimized by hybridization between DE and SA, DESA. Experimental results based on classification accuracy (in %) to measure the performance of the different classifier compared with our approach, accuracy calculated as in Eq. (4.6).

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.6}$$

Where, TP: true positive, TN: true negative, FP: false positive, FN: false negative.

4.2.5 Diseases Data Sets

Classification methods and optimization algorithms are hybridized on the following disease data sets:

(1) PID

This data set comprises patients who are pregnant females at least 21 years old and of Pima Indian heritage according to the National Institute of Diabetes and Digestive and Kidney Diseases.

Number of instances: 768

Number of attributes: 9 (including class attribute).

(2) LD

Seven attributes are included in this data set. Blood test results related to liver disorders due to alcohol consumption are indicated in the five attributes. The number of drinks per day constitutes the sixth attribute. Patient class and condition (i.e., whether the patient has the disorder or not) are presented in the seventh attribute.

Number of instances: 345

Number of attributes: 7 (including class attribute)

(3) Anemia

real data set taken from blood laboratory in AL-Anbar health directorate / Iraq [125]. Anemia is an indication of a low level of hemoglobin, which will cause a decrease in the level of oxygen transfer to the tissues of the human body. Hoque et al. [126]. A Complete Blood Cell test conducted for patients in laboratory. Anemia data is real dataset gathered from Iraqi blood laboratories to detect Anemia diseases. The anemia diagnosing identified using this information: age, gender, hemoglobin (HP), Hematocrit (HCT) and other attribute values shown in Table 4.2.

Table 4.2: Anemia dataset specifications (attributes)

Abbreviations	Explanation	Units
Age	Age of patients	Years
Gender	Gender of patients	1: male 2: female
HP	Hemoglobin	(G/dl $\times 10^6$)
RBC	Red blood cell	(Cell / microliter $\times 10^3$)
HCT	Hematocrit	(%)
MCV	Mean Corpuscular Volume	(fL)
MCH	Mean Corpuscular Hemoglobin	(Picogram)
MCHC	Mean Corpuscular Hemoglobin Concentration	(G/dl $\times 10^6$)
WBC	White blood cell	(Cell / microliter $\times 10^3$)
PLT	platelets	(Cell / microliter $\times 10^3$)

In general, Table 4.3 shows briefly description for datasets used in this study.

Table 4.3: Description of Datasets

Dataset	No. attributes	No. records
Anemia	11	803
Pima Indian diabetes (PID)	8	768
Liver disorders (LD)	9	286

(4)

4.3 RESULTS

In the context of the study three diseases datasets Pima Indian Diabetes PID and Liver Disorder from UCI repository and Anemia from AL-Anbar health provenance laboratories are tested with two trajectory algorithms SA and TS and two evolutionary algorithms GA and DE finally with proposal method DESA. Empirical results are shown in Table 4.4.

Table 4.4: Accuracy of the three diseases datasets tested by ANN and the five algorithms

	ANN	TS	SA	GA	DE	DESA
Anemia	94.14	97.01	96.77	96.64	96.52	**97.1**
Pima Indian Diabetes	78.91	80.86	80.86	82.03	80.86	**84**
Liver Disorder	75.65	78.84	78.26	77.68	78.26	**79.8**

Obviously, DESA overcome all other algorithms results, as a result, study concludes that balance between exploration and exploitation would widely be advantageous to explore search space and intensively exploit most promising solutions within local area of the search and balance between the two operations.

Statistical t test performed after 25 runs of ANN against ANN+DEASA, to test the statistical difference significant. The t-test statistics is converted into a conditional probability called p-value. The question, "If the null hypothesis is true, then what is the probability of observing the current data or data that are more extreme?" is answered by the p-value (3.7) in section **Error! Reference source not found.**. P value of null hypotheses was less than 0.0001 and the difference extremely significant, test values listed below in Table 4.5.

Table 4.5: Statistical test and of 3 medical datasets: Pima Indian Diabetes PID, Liver Disorder LD, and Anemia, between ANN and ANN + DESA

	PIMA	**LD**	**Anemia**
Mean difference	0.03684000	0.002676	0.010216
T score	14.1156	36.5694	11.8647

standard error of difference	0.005	0.002	0.001
Difference verdict	Significant	Significant	Significant

4.4 DISCUSSION

Number of hidden layers and number of neurons in every layer id critical in ANN leaning. Build suitable ANN structure by select best number of hidden layer and best number of neurons in each layer such that this structure produces best accuracy in medical dataset classification. Boosting ANN model complexity by rising number of hidden layer and number of neurons will potentially be led to overfitting. Metaheuristic algorithms normally solve ANN overfitting problem. But in the context of this study, ANN exclusively pushed for a potential high overfitting in training phase. Therefore, ANN need a metaheuristic method able to explore search space to fulfill diversification at the same time perform intensive exploit for most promising solutions in local areas.

DESA method represent an example for balance between exploration and exploitation the exploration feature of the DE algorithm and the exploitation feature of the SA algorithm are combined to reach the global optimum solution.

It is not mandatory when hybridize two metaheuristic algorithms among a list of metaheuristic algorithms to select two high accuracy ones. It is not like that. As shown in Table 4.4, and illustrate in Figure 4.4, in anemia as example we can notice accuracy of TS is more than accuracy of SA, and accuracy of GA is greater than accuracy of DE, despite that best hybridization was between DE and SA, this hybridization combine between the two algorithms perfectly and overcome all the other algorithms.

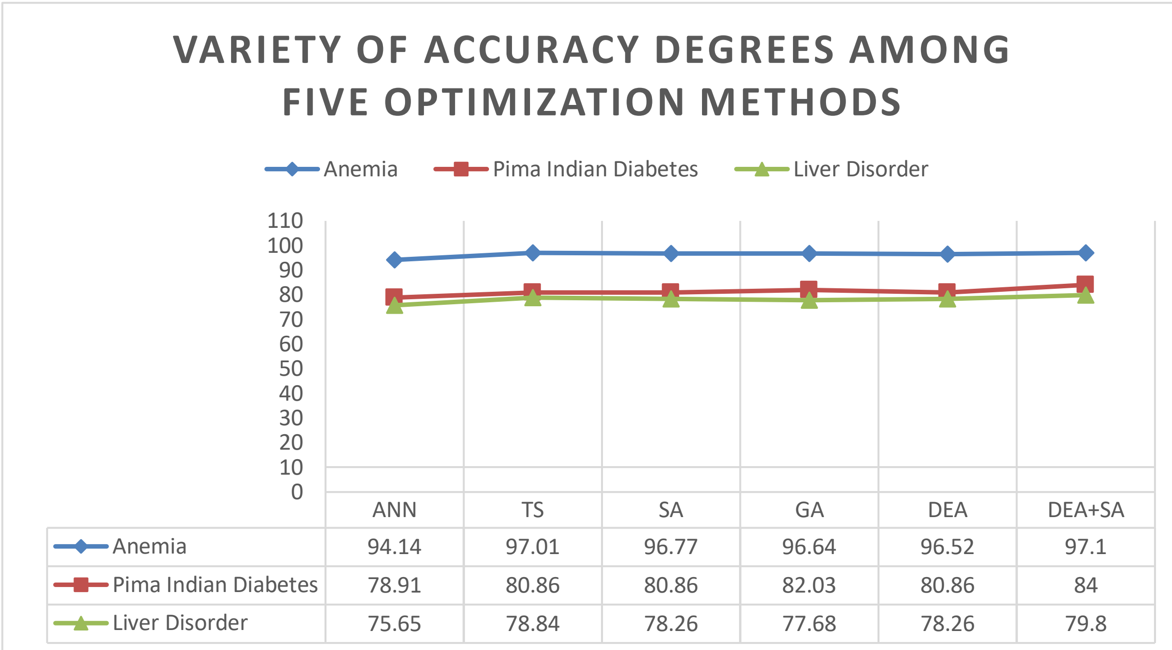


Figure 4.4: Accuracy degrees of Anemia, PID, and LD, tested with the optimization methods

All metaheuristic algorithms chosen in this study record enhancement when hybridized with ANN in spite of the improvement is low and the difference of the enhancement is not look very big but accuracy results of classification are high considering the literature and the study in chapter 3, and the reason for this is that we have already clicked on the neural network model as much as possible to find the greatest value for accuracy

4.5 CONCLUSIONS

Proposing metaheuristic method facilitates balance between exploration and exploitation to enhance medical datasets classification was the aim of this study. Number of hidden layers with number of neurons in each layer could affect ANN learning. Prepare ANN structure by select complex structure produce high accuracy can be useful to try on metaheuristic algorithms efficiency to search of global optimum. When combine two metaheuristic algorithms to formulate new better one for any reason, it is not mandatory to choose the two with high accuracy results, it is a matter of empirical tests to find convenience. Our approach can be adopted in future studies to solve various problems.

5. CONCLUSION

This chapter presents the overall conclusions of the major expansions corresponded to the work reported in this thesis. This integration comes up with mapping between research questions, research objective, and contributions conducted in this thesis. The overall conclusion of the works using standard and real-world data proposed in the context of the thesis.

In the first study, ANN is hybridized with three optimization algorithms. They are GA, PSO, and FWA. PSO is adopted to hybridize ANN with diverse layer structures with different number of metaheuristic iterations, and its parameters are tuned. Unlike most of other approaches in the literature, PSO is competitive on the five utilized medical data sets.Our approach can be adopted in future studies to solve various medical problems.

In the second study a hybridization between Differential evolution (DE) and simulating annealing (SA) to enhance artificial neural network and proposed ANN_DESA method in order to balance between exploration (diversification) and explanation (intensification), the method tested on Anemia real medical data set classification. Proposed methodology begins with choosing best ANN structure (best number of hidden layers and best number of neurons in each layer) then classification process tackle with hybridize DEA and SA to obtain batter classification accuracy. Statistical t test of ANN against ANN+DEASA performed to test the statistical difference significant, p value of null hypotheses obtained to find difference significant. Proposing metaheuristic method facilitates balance between exploration and exploitation to enhance medical datasets classification was the aim of this study. Number of hidden layers with number of neurons in each layer could affect ANN learning. Prepare ANN structure by select complex structure produce high accuracy can be useful to try on metaheuristic algorithms efficiency to search of global optimum. When combine two metaheuristic algorithms to formulate new better one for any reason, it is not mandatory to choose the two with high accuracy results, it is a matter of empirical tests to find convenience. For example, in the context of the conducted study SA produce lower result than TS, and GA overcome DE when hybrid with ANN, but after combine DE and SA in ANN+DEASA a remarkable result notice. In spite of no scientific or mathematic rules govern the implementation of metaheuristic hybridization except empirical tests, it still interesting field for research in multiple problem domains including medical problems.

REFERENCES

[1] D. R. Tobergte and S. Curtis, *Data Mining know it all*, vol. 53, no. 9. 2013.

[2] W. Frawley, G. Piatetsky-Shapiro, and C. Matheus, *Knowledge Discovery in Databases: An Overview.* AI magazine, 1992.

[3] J. Han and M. Kamber, “Data Mining : Concepts and Techniques,” *Techniques*, 2006.

[4] J. Han and M. Kamber, *Data mining: Concepts, models and techniques*, vol. 12. 2011.

[5] L. Fausett, “Fundamentals of Neural Networks,” *Igarss 2014*, no. 1, pp. 1–5, 2014.

[6] J. A. Sáez, B. Krawczyk, and M. Woźniak, “On the Influence of Class Noise in Medical Data Classification: Treatment Using Noise Filtering Methods,” *Appl. Artif. Intell.*, vol. 30, no. 6, pp. 590–609, 2016.

[7] X. Liu, H. Fu, X. Liu, and H. Fu, “PSO-Based Support Vector Machine with Cuckoo Search Technique for Clinical Disease Diagnoses,” *Sci. World J.*, vol. 2014, pp. 1–7, 2014.

[8] K. Kayaer and T. Yildirim, “Medical Diagnosis on Pima Indian Diabetes Using General Regression Neural Networks,” *Iternational Conf. Artif. Neural Networks Neural Inf. Process.*, pp. 181–184, 2003.

[9] H. N. A. Pham and E. Triantaphyllou, “An application of a new meta-heuristic for optimizing the classification accuracy when analyzing some medical datasets,” *Expert Syst. Appl.*, vol. 36, no. 5, pp. 9240–9249, 2009.

[10] H. N. A. Pham and E. Triantaphyllou, “A meta-heuristic approach for improving the accuracy in some classification algorithms,” *Comput. Oper. Res.*, vol. 38, no. 1, pp. 174–189, 2011.

[11] A. Mert, K. Niyazi, E. Bilgili, and A. Akan, “Breast Cancer Detection with Reduced Feature Set,” *Comput. Math. Methods Med.*, vol. Article ID, p. 11 pages, 2014.

[12] D. Zeng, K. Liu, S. Lai, G. Zhou, and J. Zhao, “Relation Classification via Convolutional Deep Neural Network,” no. 2011, pp. 2335–2344, 2014.

[13] C. Blum and A. Roli, “Metaheuristics in combinatorial optimization: overview and conceptual comparison,” *ACM Comput. Surv.*, vol. 35, no. 3, pp. 189–213, 2003.

[14] S. Kang and C. Isik, “Partially Connected Feedforward Neural Networks Structured by Input Types,” no. January, 2005.

[15] S. Yang and Y. Chen, “An evolutionary constructive and pruning algorithm for artificial neural networks and its prediction applications,” *Neurocomputing*, vol. 86, pp. 140–149, 2012.

[16] S. Oh and W. Pedrycz, “A new approach to self-organizing fuzzy polynomial neural networks guided by genetic optimization,” vol. 345, pp. 88–100, 2005.

[17] S. Oh, W. Pedrycz, and B. Park, “Polynomial neural networks architecture : analysis and design,” vol. 29, pp. 703–725, 2003.

[18] S. Oh, W. Pedrycz, and S. Roh, “Hybrid fuzzy set-based polynomial neural networks and their development with the aid of genetic optimization and information granulation,” vol. 9, pp. 1068–1089, 2009.

[19] B. Ludermir, “An Optimization Methodology for Neural Network,” vol. 17, no. 6, pp. 1–25, 2006.

[20] C. Zanchettin, T. B. Ludermir, and L. M. Almeida, “Hybrid training method for MLP: Optimization of architecture and training,” *IEEE Trans. Syst. Man, Cybern. Part B Cybern.*, vol. 41, no. 4, pp. 1097–1109, 2011.

[21] Bo Yuan, M. Gallagher, B. Y. B. Yuan, and M. Gallagher, “A Hybrid Approach to

Parameter Tuning in Genetic Algorithms," *2005 IEEE Congr. Evol. Comput.*, vol. 2, pp. 1096–1103, 2005.
[22] A. E. Eiben and S. K. Smit, "Parameter tuning for configuring and analyzing evolutionary algorithms," *Swarm Evol. Comput.*, vol. 1, no. 1, pp. 19–31, 2011.
[23] B. Yuan and M. Gallagher, "Combining meta-EAs and racing for difficult EA parameter tuning tasks," *Stud. Comput. Intell.*, vol. 54, pp. 121–142, 2007.
[24] S. M. Mousavi, V. Hajipour, S. T. A. Niaki, and N. Alikar, "Optimizing multi-item multi-period inventory control system with discounted cash flow and inflation: Two calibrated meta-heuristic algorithms," *Appl. Math. Model.*, vol. 37, no. 4, pp. 2241–2256, 2013.
[25] S. K. Smit, *Parameter Tuning and Scientific Testing in Evolutionary Algorithms*. 2012.
[26] S. Adam, "Steering of balance between exploration and exploitation properties of evolutionary algorithms - Mix selection," *Artifical Intell. Soft Comput.*, vol. 6114, no. Part 2, p. 13232, 2010.
[27] A. Al-Naqi, A. T. Erdogan, and T. Arslan, "Balancing exploration and exploitation in an adaptive three-dimensional cellular genetic algorithm via a probabilistic selection operator," *2010 NASA/ESA Conf. Adapt. Hardw. Syst. AHS 2010*, pp. 258–264, 2010.
[28] H. Valizadegan, R. Jin, and S. Wang, "Learning to trade off between exploration and exploitation in multiclass bandit prediction," *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discov. data Min. - KDD '11*, p. 204, 2011.
[29] M. Črepinšek, S.-H. Liu, and M. Mernik, "Exploration and exploitation in evolutionary algorithms," *ACM Comput. Surv.*, vol. 45, no. 3, pp. 1–33, 2013.
[30] M. Tokic, "Gradient Algorithms for Exploration / Exploitation Trade-Offs : Global and Local Variants," *Artif. Neural Networks Pattern Recognit.*, vol. 1, pp. 60–71, 2012.
[31] B. Yu, B. Jiao, and X. Gu, "Cooperative particle swarm optimizer based on multi-population and its application to Flow-Shop Scheduling Problem," *2008 Asia Simul. Conf. - 7th Int. Conf. Syst. Simul. Sci. Comput. ICSC 2008*, no. 4675620, pp. 1536–1542, 2008.
[32] C. H. Tan, C. K. Goh, K. C. Tan, and A. Tay, "A cooperative coevolutionary algorithm for multiobjective particle swarm optimization," *2007 IEEE Congr. Evol. Comput.*, no. June 2014, pp. 3180–3186, 2007.
[33] B. Niu, Y.-L. Zhu, and X.-X. He, "A multi-population cooperative particle swarm optimizer for neural network training," vol. 3971, pp. 570–576, 2005.
[34] B. Niu, Y. Zhu, X. He, and H. Wu, "MCPSO: A multi-swarm cooperative particle swarm optimizer," *Appl. Math. Comput.*, vol. 185, no. 2, pp. 1050–1062, 2007.
[35] D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization," *IEEE Trans. Evol. Comput.*, vol. 1, no. 1, pp. 67–82, 1997.
[36] Y. C. Ho and D. L. Pepyne, "Simple explanation of the no-free-lunch theorem and its implications," *J. Optim. Theory Appl.*, vol. 115, no. 3, pp. 549–570, 2002.
[37] I. Aljarah, H. Faris, and S. Mirjalili, "Optimizing connection weights in neural networks using the whale optimization algorithm," *Soft Comput.*, vol. 22, no. 1, pp. 1–15, 2016.
[38] I. S. Jasim, A. D. Duru, K. Shaker, B. M. Abed, and H. M. Saleh, "Evaluation and Measuring Classifiers of Diabetes Diseases," *ICET2017*.
[39] A. E. Kazdin, "Evidence-Based Treatment and Practice: New Opportunities to Bridge Clinical Research and Practice, Enhance the Knowledge Base, and Improve Patient Care," *Am. Psychol.*, vol. 63, no. 3, pp. 146–159, 2008.
[40] D. Atkins *et al.*, "Assessing applicability when comparing medical interventions: AHRQ and the Effective Health Care Program," *J. Clin. Epidemiol.*, vol. 64, no. 11, pp. 1198–1207,

2011.

[41] E. W. T. Ngai, Y. Hu, Y. H. Wong, Y. Chen, and X. Sun, “The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature,” *Decis. Support Syst.*, vol. 50, no. 3, pp. 559–569, 2011.

[42] C. Giraud-Carrier and O. Povel, “Characterising Data Mining Software,” *Intell. Data Anal.*, vol. 7, no. 3, pp. 181–192, 2003.

[43] S. Beniwal and J. Arora, “Classification and Feature Selection Techniques in Data Mining,” *Int. J. Eng. Res. Technol. data Min.*, vol. 1, no. 6, pp. 1–6, 2012.

[44] R. Caruana and A. Niculescu-Mizil, “An Empirical Comparison of Supervised Learning Algorithms,” *Proc. 23rd Int. Conf. Mach. Learn.*, pp. 161–168, 2006.

[45] S. B. Kotsiantis, “Supervised Machine Learning: A Review of Classification Techniques,” *Informatica*, vol. 31, pp. 249–268, 2007.

[46] D. R. Deepathi, K. Eswaran, S. M. Plaza, and A. S. R. Nagar, “Automatic Pattern Classification By Unsupervised Learning Using Dimensionality Reduction of Data With Mirroring,” *Neural Networks*, no. 1.

[47] R. Sathya and A. Abraham, “Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification,” *Int. J. Adv. Res. Artif. Intell.*, vol. 2, no. 2, pp. 34–38, 2013.

[48] A. Kalantari, A. Kamsin, S. Shamshirband, A. Gani, H. Alinejad-Rokny, and A. T. Chronopoulos, “Computational intelligence approaches for classification of medical data: State-of-the-art, future challenges and research directions,” *Neurocomputing*, vol. 0, pp. 1–21, 2017.

[49] A. G. Karegowda, M. A. Jayaram, and A. S. Manjunath, “Cascading K-means Clustering and K-Nearest Neighbor Classifier for Categorization of Diabetic Patients,” *Int. J. Eng. Adv. Technol.*, vol. 1, no. 3, pp. 147–151, 2012.

[50] K. Saxena, Z. Khan, and S. Singh, “Diagnosis of Diabetes Mellitus using K Nearest Neighbor Algorithm,” *Int. J. Comput. Sci. Trends Technol.*, vol. 2, no. 4, pp. 36–43, 2014.

[51] M. Behroozi and A. Sami, “A multiple-classifier framework for Parkinson’s disease detection based on various vocal tests,” *Int. J. Telemed. Appl.*, vol. 2016, 2016.

[52] C. Güzel and F. Engineering, “AWERProcedia Information Technology & Computer Science Breast Cancer Diagnosis Based on Naïve Bayes Machine Learning Classifier with KNN Missing Data Imputation,” vol. 04, pp. 401–407, 2013.

[53] L. a. Zadeh, “Fuzzy sets,” *Inf. Control*, vol. 8, no. 3, pp. 338–353, 1965.

[54] H. R. Marateb and S. Goudarzi, “A noninvasive method for coronary artery diseases diagnosis using a clinically-interpretable fuzzy rule-based system,” *J. Res. Med. Sci.*, vol. 20, no. March, pp. 214–223, 2015.

[55] J. A. Sanz, M. Galar, A. Jurio, A. Brugos, M. Pagola, and H. Bustince, “Medical diagnosis of cardiovascular diseases using an interval-valued fuzzy rule-based classification system,” *Appl. Soft Comput. J.*, vol. 20, pp. 103–111, 2014.

[56] S. Y. Ho, C. H. Hsieh, H. M. Chen, and H. L. Huang, “Interpretable gene expression classifier with an accurate and compact fuzzy rule base for microarray data analysis,” *BioSystems*, vol. 85, no. 3, pp. 165–176, 2006.

[57] A. Daemen *et al.*, “Improved modeling of clinical data with kernel methods,” *Artif. Intell. Med.*, vol. 54, no. 2, pp. 103–114, 2012.

[58] J. Shawe-Taylor and N. Cristianini, *Kernel methods for pattern analysis*. Cambridge university press, 2004.

[59] T. S. Furey, N. Cristianini, N. Duffy, D. W. Bednarski, M. Schummer, and D. Haussler, "Support vector machine classification and validation of cancer tissue samples using microarray expression data," vol. 16, no. 10, pp. 906–914, 2000.
[60] I. GUYON and J. W. S. BARNHILL, "Gene selection for cancer classification," pp. 389–422, 2009.
[61] B. Zheng, S. W. Yoon, and S. S. Lam, "Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms," *Expert Syst. Appl.*, vol. 41, no. 4 PART 1, pp. 1476–1482, 2014.
[62] M. Seera and C. P. Lim, "A hybrid intelligent system for medical data classification," *Expert Syst. Appl.*, vol. 41, no. 5, pp. 2239–2249, 2014.
[63] M. Durairaj and G. Kalaiselvi, "Prediction of Diabetes Using Back Propagation Algorithm," *Int. J. Emerg. Technol. Innov. Eng.*, vol. 1, no. 8, pp. 21–25, 2015.
[64] D. Heider *et al.*, "A Computational Approach for the Identification of Small GTPases Based on Preprocessed Amino Acid Sequences," *Technol. Cancer Res. Treat.* , vol. 8, no. 5, pp. 333–341, 2009.
[65] T. Jayalakshmi and a Santhakumaran, "Statistical normalization and back propagation for classification," *Int. J. Comput. …*, vol. 3, no. 1, pp. 1–5, 2011.
[66] T. Kiyan and T. Yildirim, "Breast cancer diagnosis using statistical neural networks," *IU-Journal Electr. Electron. Eng.*, vol. 4, no. 2, pp. 1149–1153, 2011.
[67] P. Uskaner, "HABERMAN'S SURVIVAL IN ARTIFICIAL NEURAL NETWORKS EECS- 589 İntroduction to Artificial Neural Network Pınar Uskaner 22.05.2014," 2014.
[68] P. Seidel, A. Seidel, and O. Herbarth, "Multilayer perceptron tumour diagnosis based on chromatography analysis of urinary nucleosides," *Neural Networks*, vol. 20, no. 5, pp. 646–651, 2007.
[69] H. Yan, Y. Jiang, J. Zheng, C. Peng, and Q. Li, "A multilayer perceptron-based medical decision support system for heart disease diagnosis," *Expert Syst. Appl.*, vol. 30, no. 2, pp. 272–281, 2006.
[70] M. R. Amiryousefi, M. Zarei, M. Azizi, and M. Mohebbi, "Modelling some physical characteristics of pomegranate (Punica granatum L.) fruit during ripening using artificial neural network," *J. Agric. Sci. Technol.*, vol. 14, no. 4, pp. 857–867, 2012.
[71] I. Boussaïd, J. Lepagnot, and P. Siarry, "Infor mation Sciences A survey on optimization metaheuristics," vol. 237, pp. 82–117, 2013.
[72] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of State Calculations by Fast Computing Machines," vol. 1087, 1953.
[73] H. H. Örkcü, M. Đ. Doğan, and M. Örkcü, "A hybrid applied optimization algorithm for training multi-layer neural networks in data classification A Hybrid Applied Optimization Algorithm for Training Multi-Layer Neural Networks in Data Classification," no. January 2015, 2017.
[74] M. Alweshah, "Firefly Algorithm with Artificial Neural Network for Time Series Problems," vol. 7, no. 19, pp. 3978–3982, 2014.
[75] F. Glover, "Future paths for integer programming and links to artificial intelligence," *Comput. Oper. Res.*, vol. 13, no. 5, pp. 533–549, 1986.
[76] E.-G. Talbi, *METAHEURISTICS FROM DESIGN TO IMPLEMENTATION*. .
[77] J. H. HOLLAND, "Adaptation in Natural and Artificial Systems," *Univ. Michigan Press*, 1975.
[78] M. Mitchell, "An introduction to genetic algorithms," *Comput. Math. with Appl.*, vol. 32,

no. 6, p. 133, 1996.
[79] D. E. Goldberg and J. H. Holland, "Genetic Algorithms and Machine Learning," *Mach. Learn.*, vol. 3, no. 2, pp. 95–99, 1988.
[80] S. Mizuta, T. Sato, D. Lao, M. Ikeda, and T. Shimizu, "Structure Design of Neural Networks Using Genetic Algorithms," *Complex Syst.*, vol. 13, no. 2, pp. 161–175, 2001.
[81] H. H. Örkcü, M. Ð. Doğan, and M. Örkcü, "A Hybrid Applied Optimization Algorithm for Training Multi-Layer Neural Networks in Data Classification," vol. 28, no. 1, pp. 115–132, 2015.
[82] I. Salman, "Impact of Metaheuristic Iteration on Artificial Neural," 2018.
[83] B. M. Abed, K. Shaker, H. A. Jalab, H. Shaker, A. M. Mansoor, and A. F. Alwan, "Breast Cancer Diagnosis A Hybrid Classification Algorithm Approach for Breast Cancer Diagnosis," no. April 2017, 2016.
[84] M. A. Jabbar, B. L. Deekshatulu, and P. Chandra, "Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm," *Procedia Technol.*, vol. 10, pp. 85–94, 2013.
[85] D. Avci and A. Dogantekin, "An Expert Diagnosis System for Parkinson Disease Based on Genetic Algorithm-Wavelet Kernel-Extreme Learning Machine," *Parkinsons. Dis.*, vol. 2016, 2016.
[86] L. Shen *et al.*, "Evolving support vector machines using fruit fly optimization for medical data classification," *Knowledge-Based Syst.*, vol. 96, no. January, pp. 61–75, 2016.
[87] D. K. Choubey and S. Paul, "Ga_J48graft Dt: A hybrid intelligent system for diabetes disease diagnosis," *Int. J. Bio-Science Bio-Technology*, vol. 7, no. 5, pp. 135–150, 2015.
[88] M. M. Mafarja, D. Eleyan, I. Jaber, A. Hammouri, and S. Mirjalili, "Binary Dragonfly Algorithm for Feature Selection," in *2017 International Conference on New Trends in Computing Sciences (ICTCS)*, 2017, pp. 12–17.
[89] N. S. Jaddi and S. Abdullah, "Hybrid of genetic algorithm and great deluge algorithm for rough set attribute reduction," *Turkish J. Electr. Eng. Comput. Sci.*, vol. 21, no. 6, pp. 1737–1750, 2013.
[90] Z. Zainuddin, K. H. Lai, and P. Ong, "An enhanced harmony search based algorithm for feature selection : Applications in epileptic seizure detection and prediction R," *Comput. Electr. Eng.*, vol. 53, pp. 143–162, 2016.
[91] R. Storn and K. PRICE, "Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces," pp. 341–359, 1997.
[92] H. T. Tike Thein and K. M. Mo Tun, "An Approach for Breast Cancer Diagnosis Classification Using Neural Network," *Adv. Comput. An Int. J.*, vol. 6, no. 1, pp. 1–11, 2015.
[93] O. S. Soliman and E. Aboelhamd, "Classification of Breast Cancer using Differential Evolution and LeastSquares ABSTRACT :," vol. 3, no. 2, pp. 155–161, 2014.
[94] I. De Falco, "Differential Evolution for automatic rule extraction from medical databases," *Appl. Soft Comput. J.*, vol. 13, no. 2, pp. 1265–1283, 2013.
[95] H. A. Abbass, "An evolutionary artificial neural networks approach for breast cancer diagnosis," *Artif Intell Med*, vol. 25, no. 3, pp. 265–281, 2002.
[96] J. Kennedy and R. Eberhart, "Particle swarm optimization," *Neural Networks, 1995. Proceedings., IEEE Int. Conf.*, vol. 4, pp. 1942–1948 vol.4, 1995.
[97] F. Ardjani, K. Sadouni, and M. Benyettou, "Optimization of SVM multiclass by particle swarm (PSO-SVM)," *2010 2nd Int. Work. Database Technol. Appl. DBTA2010 - Proc.*, no. December, pp. 32–38, 2010.

[98] M. G. Feshki and O. S. Shijani, “Improving the Heart Disease Diagnosis by Evolutionary Algorithm of PSO and Feed Forward Neural Network,” pp. 48–53, 2016.
[99] R. K. Dutta, N. K. Karmakar, and T. Si, “Artificial Neural Network Training using Fireworks Algorithm in Medical Data Mining,” vol. 137, no. 1, pp. 1–5, 2016.
[100] P. Shrivastava, A. Shukla, P. Vepakomma, N. Bhansali, and K. Verma, “A survey of nature-inspired algorithms for feature selection to identify Parkinson’s disease,” *Comput. Methods Programs Biomed.*, vol. 139, pp. 171–179, 2017.
[101] M. Fatima and M. Pasha, “Survey of Machine Learning Algorithms for Disease Diagnostic,” pp. 1–16, 2017.
[102] P. Suganya and C. P. Sumathi, “A novel metaheuristic data mining algorithm for the detection and classification of Parkinson Disease,” *Indian J. Sci. Technol.*, vol. 8, no. 14, 2015.
[103] P. Tavakkoli, D. M. Souran, S. Tavakkoli, M. Hatamian, A. Mehrabian, and V. E. Balas, “Classification of the liver disorders data using Multi-Layer adaptive Neuro-Fuzzy inference system,” *6th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT 2015*, pp. 13–16, 2016.
[104] S. Mandal and I. Banerjee, “Cancer Classification Using Neural Network,” vol. 3, no. 7, pp. 172–178, 2015.
[105] T. Desell, S. Clachar, J. Higgins, and B. Wild, “Evolving Deep Recurrent Neural Networks using Ant Colony Optimization,” *Evol. Comput. Comb. Optim.*, vol. 4446, no. 12, pp. 154–165, 2007.
[106] C. Blum, “Université Libre de Bruxelles Training feed-forward neural networks with ant colony optimization : An application to pattern classification IRIDIA – Technical Report Series,” no. June, 2014.
[107] M. Lichman, “UCI Machine Learning Repository; University of California, School of Information and Computer Science: Irvine, CA, USA, 2013. Available online:,” 2013. [Online]. Available: http://archive.ics.uci.edu/ml. [Accessed: 13-Mar-2016].
[108] K. V. S. R. P. Varma, A. A. Rao, T. Sita Maha Lakshmi, and P. V. Nageswara Rao, “A computational intelligence approach for a better diagnosis of diabetic patients,” *Comput. Electr. Eng.*, vol. 40, no. 5, pp. 1758–1765, 2014.
[109] M. Maddouri and M. Elloumi, “A data mining approach based on machine learning techniques to classify biological sequences,” vol. 15, pp. 217–219, 2002.
[110] J. Luo, M. Wu, D. Gopukumar, and Y. Zhao, “Big Data Application in Biomedical Research and Health Care : A Literature Review,” *Biomed Inf. Insights*, vol. 8, pp. 1–10, 2016.
[111] M. R. G. S. C. P. M. Pardalos, “Classification and Characterization of Gene Expression Data with Generalized Eigenvalues,” pp. 533–545, 2009.
[112] Y. Tan, “Fireworks Algorithm,” pp. 247–262, 2015.
[113] Y. Yuan *et al.*, “Assessing the clinical utility of cancer genomic and proteomic data across tumor types,” *Nat. Biotechnol.*, vol. 32, p. 644, Jun. 2014.
[114] V. Cestarelli, G. Fiscon, G. Felici, P. Bertolazzi, and E. Weitschek, “CAMUR: Knowledge extraction from RNA-seq cancer data through equivalent classification rules,” *Bioinformatics*, vol. 32, no. 5, pp. 697–704, Mar. 2016.
[115] A. A. Afifi and S. P. Azen, *Statistical Analysis: A Computer Oriented Approach*. Elsevier Science, 2014.
[116] X. Yang, S. Deb, and S. Fong, “Metaheuristic Algorithms : Optimal Balance of Intensification and Diversification Metaheuristic Algorithms : Optimal Balance of

Intensification and Diversification," no. September 2014, 2013.
[117] E. H. Elshami and A. M. . Alhalees, "Automated Diagnosis of Thalassemia Based on DataMining Classifiers," no. June 2012, 2015.
[118] A. Yilmaz, "A FUZZY EXPERT SYSTEM DESIGN FOR," no. October, 2013.
[119] S. A. Sanap, M. Nagori, and V. Kshirsagar, "Classification of Anemia Using Data Mining Techniques BT - Swarm, Evolutionary, and Memetic Computing," 2011, pp. 113–121.
[120] N. Amin and A. Habib, "Comparison of Different Classification Techniques Using WEKA for Hematological Data," no. 3, pp. 55–61, 2015.
[121] F. Fagan and J. H. Van Vuuren, "A unification of the prevalent views on exploitation , exploration , intensification and diversification," vol. 2, no. 3, pp. 294–327, 2013.
[122] H. Makas and N. Yumus, "Balancing exploration and exploitation by using sequential execution cooperation between artificial bee colony and migrating birds optimization algorithms," pp. 4935–4956, 2016.
[123] V. Kachitvichyanukul, "Comparison of Three Evolutionary Algorithms :," vol. 11, no. 3, pp. 215–223, 2012.
[124] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, "Optimization by Simulated Annealing," vol. 220, no. 4598, pp. 671–680, 1983.
[125] "BAnemia." .
[126] M. Hoque M and M. Kader SB, "Risk factors for anaemia in pregnancy in rural KwaZulu-Natal , South Africa : Implication for health education and health promotion," vol. 51, no. 1, pp. 68–72, 2009.

APPENDIX A

[Appendix title]

First paragraph.

A.1 APPENDIX SECTION

Appendix section’s first paragraph.

Second paragraph.

A.1.1 Appendix subsection

This is a subsection (level-3 division) of appendix A.