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. Dr. Khalid SHAKER	Prof. Dr. 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 UCAN	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

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

Assoc. Prof. Dr. Oğuz BAYAT

Director

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

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.

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: Prof. 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.

Besides my advisor, I would like to thank the rest of my thesis committee for their encouragement, insightful comments.

I would like to express my very great appreciation to Altinbas University – Istanbul – Turkey

My sincere thanks also go to the Diyala University for giving me this scholarship and for their financial support which enabled me to get my PhD

I would like to thank my family: my parents for supporting me spiritually throughout my life, to my wife my children my brothers and sisters their support and encouragement throughout my study

Last but not the least, I would like to thank my friends who give me constructive comments and warm encouragement

ABSTRACT

METAHEURISTIC ALGORITHMS TO ENHANCE ARTIFITIAL NEURAL NETWORK FOR MEDICAL DATA CLASSIFICATION

Ihsan Salman Jasim Al Gburi

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

Supervisor: Prof. Dr. Osman Nuri Ucan

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

Date: September 2018

Pages: 103

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.

In the first chapter we discuss a background introduction and in the second chapter a literature survey about the approaches applied on the problem.

The third chapter of this thesis discusses the effect of metaheuristic iteration on ANN structure. The novelty of the proposed work shown through improving the impact of metaheuristic algorithm iteration on ANN structure. ANN is enhanced using separate three metaheuristic algorithms (particle swarm optimization PSO, genetic algorithm GA, and fireworks algorithm FW). The proposed models are tested on five standard medical benchmarks and one big-data medical dataset. The proposed study is successfully implemented, and remarkable results are obtained. Furthermore, the no-free-lunch theorem (NFLT) is verified in the study’s context, that is, no algorithm is universal for all problem domains.

The forth chapter of this thesis investigates the tradeoff between exploration and exploitation when obtaining the optimal global solution which represent best accuracy of medical data classification. 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 named differential evolution algorithm DE and simulated annealing SA are combined to formulate a new and improved algorithm DESA for considered problem domain. However, selecting the highly accurate two algorithms is not mandatory; instead, empirical tests can be performed for convenience.

Originality of proposed method is combining between DE as evolutionary metaheuristic algorithm and SA as trajectory algorithm to provide balance between exploration and exploitation to explore search space widely for global solutions and intensively exploited local solutions. DESA method compared with tow evolutionary which are GA and DE, and with two trajectory which are SA and Tabu search TS. Proposed method DESA is implemented successfully, and better results obtained.

Keywords: Metaheuristic algorithms, data mining, Medical datasets, Classification, ANN, Optimization

TABLE OF CONTENTS

Pages

LIST OF FIGURES .... **xiv**

LIST OF ABBREVIATIONS .... **xvi**

1. INTRODUCTION **1**

1.1 SIGNIFICANT OF THE THESIS 3

1.2 PROBLEM STATEMENT 3

1.3 RESEARCH QUESTIONS 6

1.4 RESEARCH OBJECTIVES 7

1.5 RESEARCH SCOPE 9

1.6 OVERVIEW OF THE THESIS 10

2. LITERATURE SURVEY **11**

2.1 INTRODUCTIONON 11

2.2 CLASSIFICATION PROBLEM DESCRIPTION 11

2.3 APPROACHES APPLIED TO MEDICAL PROBLEMS 14

2.3.1 Classification Approaches 16

2.3.1.1 Lazy Learner K-nearest Neighbor KNN 16

2.3.1.2 Fuzzy Sets 17

2.3.1.3 Kernel Methods SVM 18

2.3.1.4 Artificial Neural Network 18

2.3.2 Metaheuristic Approaches 22

2.3.2.1 Single Based Solution 23

2.3.2.2 Population Based Solution 25

2.4 MAIN FINDINGS FROM LITERATURE REVIEW 30

3. IMPACT OF METAHEURISTIC ITERATION ON ANN STRUCTURE **33**

3.1 INTRODUCTION 33

3.2 MATERIALS AND METHODS 35

3.2.1 Artificial Neural Network ANN 35
3.2.2 Metaheuristic Algorithms 36
3.2.2.1 Genetic Algorithm GA 36
3.2.2.2 Particle Swarm Optimization PSO 37
3.2.2.3 Fireworks algorithm FW 40
3.2.3 Proposed Work 42
3.2.4 Datasets 43
3.2.5 The Strategy for solving the problem 45
3.3 RESULTS OF THE STUDY 46
3.3.1 First Experiment 46
3.3.2 Results of Experiment 1 47
3.3.3 Experiment 2 51
3.3.4 Experiment 2 results 51
3.3.5 RNA-Seq big data test using hybridized Algorithms 55
3.4 DISCUSSION 56
3.5 CONCLUSIONS 59
4. EFFECT OF DIVERSIFICATION AND INTENSIFICATION TRADE-OFF IN ANEMIA MEDICAL DATA CLASSIFICATION 61
4.1 INTRODUCTION 61
4.2 METHODS AND MATERIALS 62
4.2.1 Artificial neural network ANN 62
4.2.2 Differential evolution DE 63
4.2.3 Simulated Annealing SA 65
4.2.4 Proposed Work 68
4.2.5 Diseases Datasets 71
4.3 RESULTS 73
4.4 DISCUSSION 74
4.5 CONCLUSIONS 75

5. CONCLUSION 76

REFERENCES...... 78

LIST OF TABLES

Pages

Table 1.1: A descriptive mapping between research questions, research objectives and contributions in this study 8

Table 2.1: Strengths and limitations of the classification methods 20

Table 3.1: Medical datasets specification....44

Table 3.2: Five medical benchmarks classification by hybridize two ANN hidden-layers with 500 iteration metaheuristic algorithms....47

Table 3.3: Five medical benchmarks classification by hybridize two ANN hidden-layers with 1000 iteration metaheuristic algorithms....47

Table 3.4: Five medical benchmarks classification by hybridize three ANN hidden-layers with 500 iteration metaheuristic algorithms....48

Table 3.5: Five medical benchmarks classification by hybridize three ANN hidden-layers with 1000 iteration metaheuristic algorithms....48

Table 3.6: Results of experiment 2, PSO algorithm for 1000 iterations to enhance 3 and 4 hidden-layer ANN 52

Table 3.7: ANN and ANN + PSO Statistical test 55

Table 3.8: RNA-Seq accuracy results from ANN and the three algorithms 56

Table 3.9: ANN and the three algorithms statistical test of RNA-Seq 56

Table 3.10: Compare proposal with the state-of-the-art results 58

Table 4.1: Parameter setting for the proposed work 70

Table 4.1: Anemia dataset specifications (attributes) 72

Table 4.3: Study benchmarks Description 72

Table 4.2: Accuracy of the three diseases datasets tested by ANN and the five algorithms 73

Table 4.5: Statistical test of the three medical datasets PID, LD, and anemia between ANN and ANN + DESA 73

LIST OF FIGURES

Pages

Figure 1.1: The evolution of database system technology. 1

Figure 1.2: Data mining as a confluence of multiple domains 2

Figure 2.1: Classification of learning systems..13

Figure 2.2: A summary of methods applied for classification problem.....................................15

Figure 3.1: Genetic algorithm (GA) 37

Figure 3.2: Particle swarm optimization PSO 39

Figure 3.3: Fireworks approach flowchart (FW) 41

Figure 3.4: The proposal methodology: ANN+GA, ANN+PSO, ANN+FW. 42

Figure 3.5: Illustration of accuracy comparing in 1000 iteration for 3 optimization algorithms, GA, PSO, and FW hybridize with 3 hidden layer ANN tested on the medical data sets: (1) PID; (2) WDBC; (3) LD; (4) HSS; (5) PD 51

Figure 3.6: PSO performances using 1000 repeated iterations and 200 population swarms with four hidden layers, tested on the medical data sets: (1) PID; (2) WDBC; (3) LD; (4) HSS; (5) PD 54

Figure 3.7: Experiment 2 enhancement...59

Figure 4.1: Crossover process for 8 parameters 64

Figure 4.2: DE flowchart..65

Figure 4.3: SA strategy to jump local minima 66

Figure 4.4: Simulated annealing algorithm 67

Figure 4.5: Hybridization between DEA and SA 68

Figure 4.6: Proposal methodology 69

Figure 4.7: Accuracy degrees of anemia, PID, and LDs under testing with optimization methods 75

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.

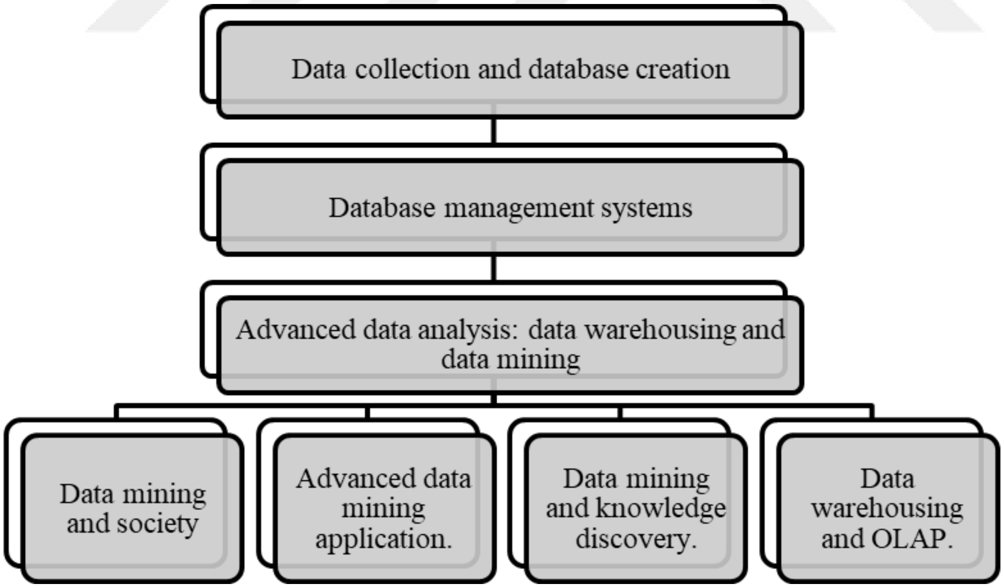


Figure 1.1: The evolution of database system technology Han et al. [3].

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.

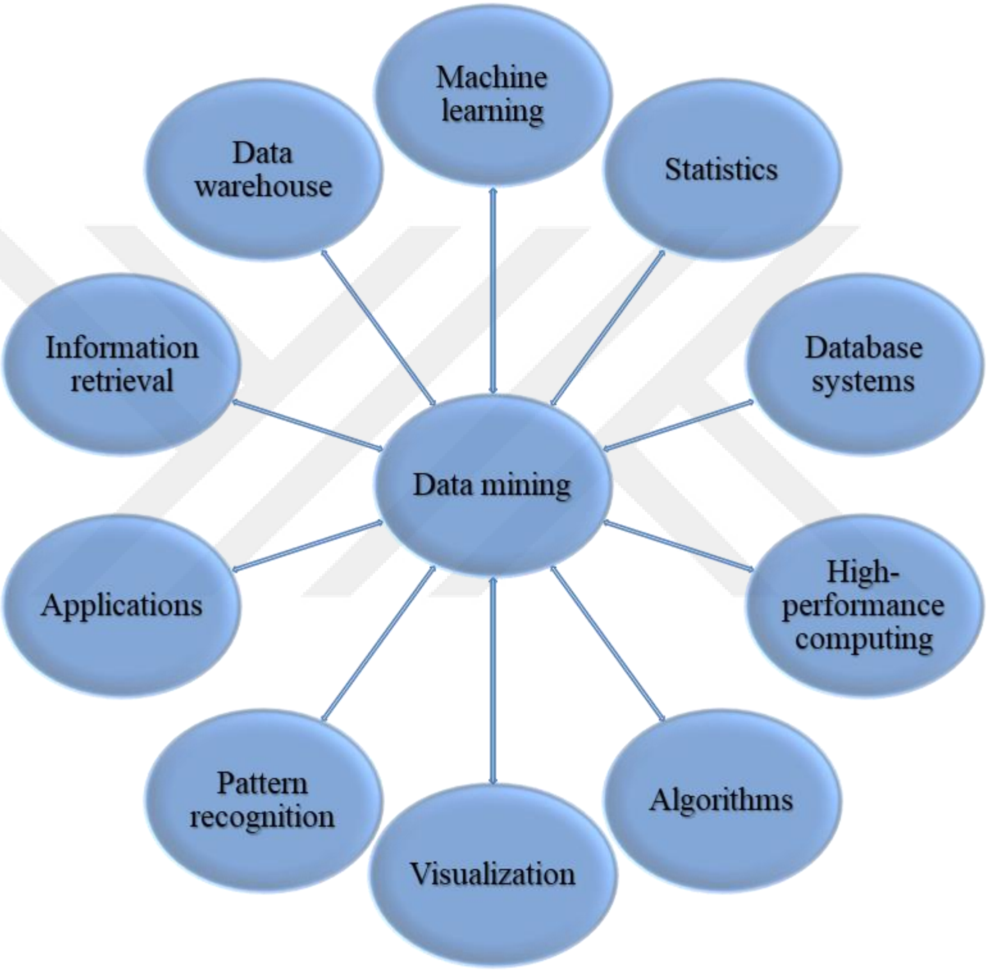


Figure 1.2: Data mining as a confluence of multiple domains Han et al. [3].

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.

The ANN is considered an imitation of the 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.

Every ANN is organized in numerous connections of the simple possessing units which named as nodes or neurons as declared in Fausett [5]. The ANN may include on one or more nodes layers between the input and output layers. These interlayers are named as hidden layers and they are connected completely or partially as mentioned by Fausett [5]. The connection which is existed between two nodes include particular 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. The automatic medical diagnosing system is possible to be extremely advantageous by carrying the whole 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 ANNs technique by the ability of nonlinearity and learning, ANNs are established at anywhere when there are prediction or classification problem. Sanggil & Isik applied many algorithms in order to tune the ANNs weights to train the model of ANN [14]. In this case, the most important element in an ANN success is the selection of ANN structure which necessitates the determination of the proper number of layers and the nodes number that existed 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, the ANN structure and weights optimization 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 enhance the neural network structure and weights instead of optimizing the weights or structure in isolate. It must be mentioned that there are only a few number of researchers who tried to enhance 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 consisted 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 considered a significant 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 factors which must be tuned previously. Through the process of solving specific problem, the parameter tuning can enhance algorithm performance (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.

A capable algorithm is needed for the optimization strategy in order to lead the optimization process to the global optimum solution (Słowik [26]). A balance of exploitation and exploration strategy in search space helps to achieve the global optimum (al-Naqi et al. [27]; Valizadegan et al. [28]). Exploration (diversification) can be explained as the ability of testing many regions in the space of problem to detect the good optimum, hopefully the global one whereas the exploitation (intensification) can be summarized as the ability of concentrating the search around a promising candidate solution to detect 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, the solution which falls in the optimality region is searched by any optimization algorithm.

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 goal 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 are described in Table 1.1.

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

Chapter	Research question	Research objective	Contribution
Third chapter	• Does changing in ANN structure with optimization of the weights 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 diversification and intensification strategy of the algorithm to improve the search process quality?	• To hybridize between evolutionary and trajectory metaheuristic method and trade-off between exploration and exploitation using modified differential evaluation algorithm DEA hybridized with simulating	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?	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 ANN model quality 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 3.3.4.

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 process between ANN structural change and metaheuristic algorithms iterations. according to that, diversity of metaheuristic iterations with different ANN structure are studied. Proposed methodology tested with big data classification that is contains thousands of attributes. A comparison conducted between methods results and further techniques in the state-of-the-art. 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 presents an overview of medical classification problems and a review of the techniques for solving these problems that have been offered in the literature until the date of performing this study. 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 in order 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 perceptions 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 (ith 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 the prediction purpose. In general, rules includes a group of conditional attributes and the attributes of class as a decision attribute. The unseen instances are classified into different classes by using the rules.

Data classification is a two-step process where in the first step, a classifier describes a determined set of data concepts. This is the training phase in which 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 ratio 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.

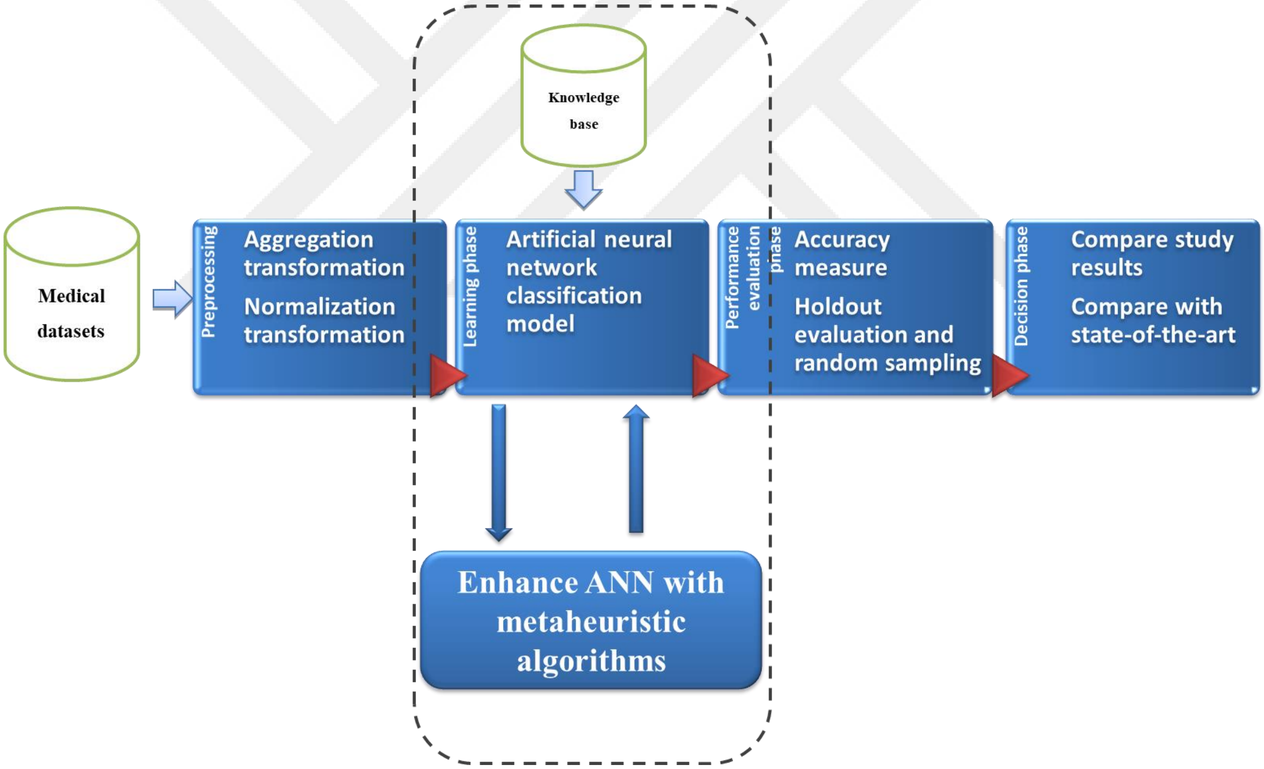


Figure 2.1: Classification of learning systems

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 deliver an overview of the techniques which are available currently. 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.

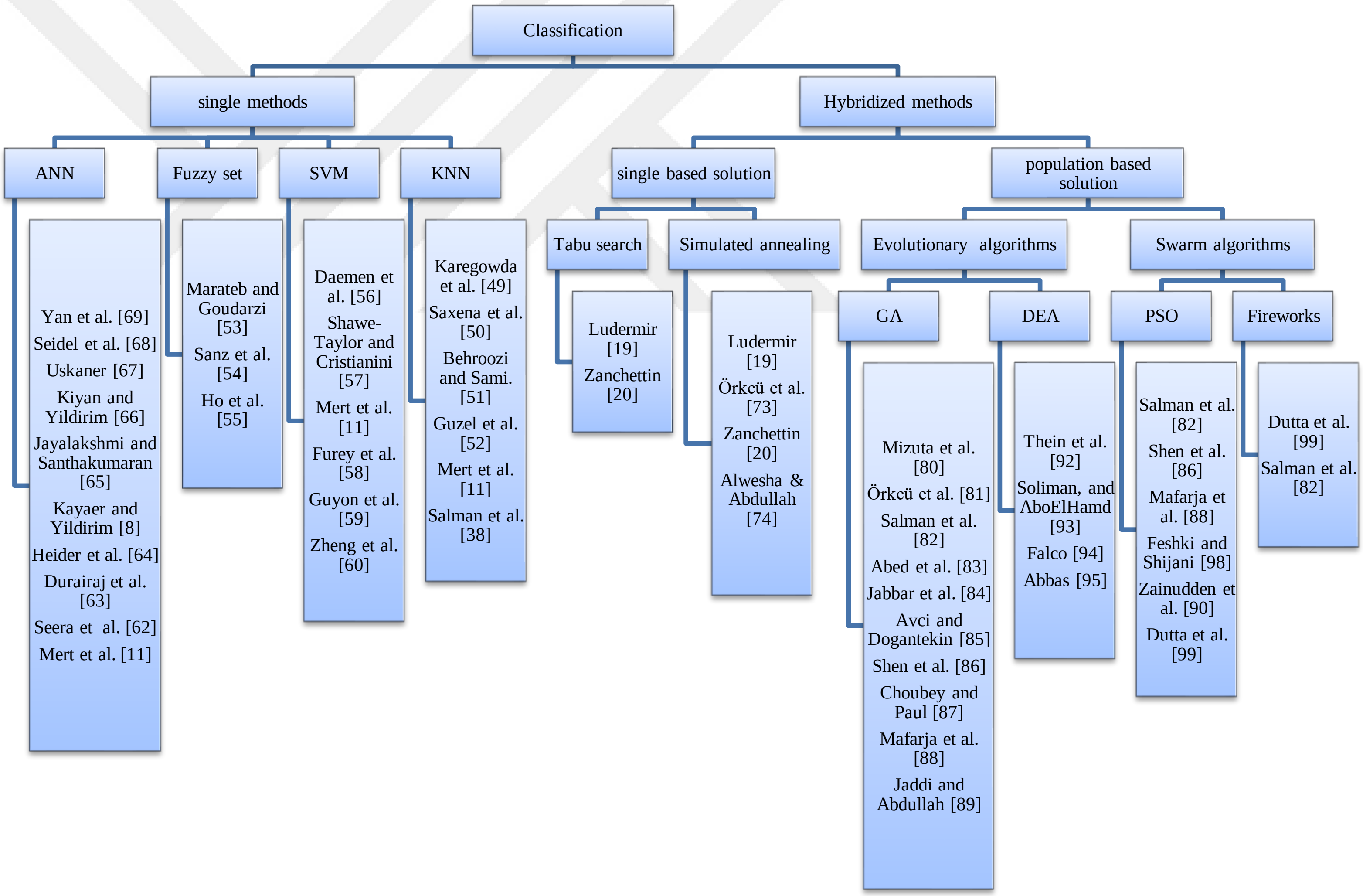


Figure 2.2: A summary of methods applied for classification problem

2.3.1 Classification Approaches

There are number of machine learning models used for classification, in the medical data mining context, classification approaches 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 relies 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.

There are multiple ways which are used to find the distance and the most important one is Euclidian's distance method that shown below:

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

This algorithm is usually used when there is new online example presents or new set of examples and must be classified in accordance with the previous data with output class.

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

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

Pima Indian Diabetes classified in spiral spinning approach 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 may be the popular vote from the whole 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 method, the membership function is distinct by the truth value of a fuzzy propositional function Zadeh 1965 [53]. According to this approach elements include 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 a new methodology in order to diagnose the cardiovascular disease during the next ten years by organizing classification systems based on fuzzy rules with interval-valued fuzzy in order to construct a classifier which able to deal with issues of identifying the risks for the patients who 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 technique is considered one of the best kernel techniques which characterizes by 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. They focused in their analysis on classifying the tissues samples and investigate the data for the mislabeled tissues results.

Guyon et al. [60] addressed the problem of assortment of a small subset of genes from the data of gene expression which registered on DNA micro-arrays by the use of training samples which available from the 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 uses 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 utilized widely for classification and rule generation. Between the vast numbers of ANN techniques which are used for classification, the simplest one is that based on a fixed number of layers and nodes. 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. The weights are adopted in each iteration by the use of new data of training data sets. Each teaching step is started by enforcing the 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 output of network and preferred 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 uses Wisconsin diagnostic breast cancer as problem domain with thirty features.

Seera et al. [62] suggested a hybrid intelligent system which consists of Regression Tree and fuzzy min-mx neural network and the Random Forest model as Classification and. The suggested hybrid system goals intensively exploit the constituent model’s advantages and, at the same time, improve their limitations, able on learning incrementally by the use of Fuzzy neural network Min–Max, use Regression Tree and Classification to clarify its expected outputs and accomplish high classification performances using 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 the structures of Neural Network can be used effectively in order to find better results. The classification accuracy of BPN with LM gotten 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] employee MLP multi-layer perceptron neural network to diagnose Haberman Survival dataset with parameters tuning 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 explained 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		• Lack of clarity in the results • Support vector machine is computational inefficiency to finding

	- Support vector machine can be easily extended to perform numerical calculations. - Support vector machine is very useful for general pattern recognition, regression and classification.	an approximate minimum enclosing 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.

The literature consists of many different types of ANN which have been proposed. 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 comprising 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 nodes number 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 involves 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 will outline single-solution also named trajectory metaheuristics methods. Unlike population-based metaheuristics, they jump with a single initial solution and start moving away from the starting point, drawing a trajectory in the search space. Some of them is ‘‘intelligent’’ evolution of local search approaches. Trajectory methods are mostly including the simulated annealing approach, the variable neighborhood search, the GRASP method, the guided local search, the Tabu search, the iterated local search, besides their variations.

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.

The general form for trajectory methods can be formulated as follows:

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

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], which they aid to permit the solver to "explore" more of the possible space of solutions through some trades that do not lower the range are recognized when Blum and Roli [13]. By use the conditions in Eq. (2.3), similar "bad" trades are certified

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

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

In following some recent related simulated annealing literature proposed in the medical data classification area

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-GSA (hybrid intelligent model between simulated Annealing and genetic Algorithm and) in order to train artificial neural networks (ANN), system aims to alleviate the genetic and simulated annealing algorithms limitations and exploit their advantages. Hybrid-GSA tested with three UCI benchmarks WDBC, PID, and LD.

Zanchettin [20], Proposed hybrid system depends on the incorporation of the replicated 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 suggesting a hybrid technique join between firefly algorithm and simulated annealing SA. Inside the firefly algorithm SA controls the randomness step whereas optimizing the standard PNN model weights. Then, the work is extended by the use of Lévy flight inside the firefly algorithm in order to explore the search space in enhance the the PNN performance and get better classification accuracy.

2. Tabu search TS

This algorithm was first introduce in 1986 by Glover [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 memory human being. It proceeds 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 pretend 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 numerous 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 named Evolutionary Algorithms (EAs)) Boussaïd et al. [71], are the spaces of genetic algorithms, genetic programming, evolutionary programming and differential evolutionary. 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 operation of crossover in accordance with crossover probability and to methodology (single point or multi point crossover).

4- Perform operation of mutation in accordance with 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 structure of network and the optimum parameter set by proposing genetic algorithms (GAs) for designing and training neural networks.

Örkcü et al. [73], develop hybrid-GSA (hybrid intelligent model between simulated Annealing and genetic Algorithm and) in order to train artificial neural networks (ANN), system aims to alleviate the genetic and simulated annealing algorithms limitations and exploit their advantages. Hybrid-GSA tested with three UCI benchmarks WDBC, PID, and LD.

Salman et al. [81] 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. [82] 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. [83] combine KNN with GA to enhance the data set of heart disease accuracy 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 [84] 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. [85] 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 [86] 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. [87] proposed wrapper-feature selection depending on the 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 [88] 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. [89] suggested a feature selection approach based on harmony algorithm to enhance the 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 [90]. 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. [91], improve performance of ANN (which always being trapped at the local minima and suffer from slow intersection), by using differential evaluation algorithm (DEA), so as to conclude the optimal or near optimal value for ANN factors, and in order to solve problems on DE method including lower classification accuracy and longer training time. The suggested system performs the island-based training technique to be less training time and better accuracy by the use of the analysis between two different migration topologies.

Soliman, and AboElHamd [92] 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 [93] 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 [94] 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 executing/ performing definite 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, planned, and autonomous algorithm Kennedy and Eberhart [95]. PSO is a regionalized population-based evolutionary approach Ardjani et al. [96].

Basically, the algorithm consists of set of particles, this set called swarm (solution symbols population in the space of search). 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 (solution symbols population) are moved to a solution within the space of search. More details found in section 3.2.2.2. in the following some related work:

Salman et al. [81] 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. [85] 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. [87] 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 [97] 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. [89] suggested a feature selection approach based on harmony algorithm in order to enhance the 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. [98], use improved firework, LM, and PSO that included an ANN model to classify five benchmark data sets from UCI repository the University of California, Irvine.

- Fireworks Algorithm

FW 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 the fireworks as illustrated in Tan [99] and Dutta et al. [98].

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

Dutta et al. [98], use improved firework, LM, and PSO that included an ANN model to classify five benchmark data sets from UCI repository the University of California, Irvine.

Salman et al. [81] 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 methods related to the classification of medical data. 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 classification area of medical data 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].

- According to 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 enhance the structure and weights of the ANN to find most appropriate model for solving classification problems.

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

3. Balancing diversification and intensification or trade-off between exploration and exploitation to develop a convergence strategy in the method to deliver a good balance between exploration and exploitation during the search process.

In this research, the aim or 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. IMPACT OF METAHEURISTIC ITERATION ON ANN STRUCTURE

3.1 INTRODUCTION

Classification of Medical data is 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 [92]. Scientists have implemented various data-mining approaches to diagnose and treat diverse diseases, like diabetes Dutta et al. [98], liver disorder Tavakkoli et al. [103], Parkinson's Shrivastava et al. 2016 [100], and cancer Örkcü et al. [104]. ANN is extensively adopted in the diseases prediction and diseases classification. Mandal and Banerjee [105] used ANN with appropriate backpropagation training, and ANN exhibits improved its efficiency by improve accuracy.

Desell et al. [106] modified the structure of deep repeated neural networks by the use of metaheuristic optimization algorithm called the ant colony. The ant colony optimization algorithm is used to train the set of weights of neural networks to optimize continuous parameter. Heider et al. [64] introduced an artificial neural network cluster that included four sub-family networks to designate a minor GTPase to a sub-family and a mesh network to classify minor GTPases. Mizuta et al. [80] aimed to generate the optimum parameter set and the best network structure by proposing genetic algorithms (GAs) for designing and training neural networks. As well as, they introduced the fitness function that relied on the errors of output and the easiness in the structure of network Blum and Socha solved discrete optimization problems via an ant colony optimization approach [107].

Örkcü et al. utilized the benefits of simulated annealing and genetic approaches in order to alleviate their limitations using an intelligent hybrid model that hybrid simulated annealing and GA aimed at training ANNs. Three benchmarks, namely, WDBC, PID, and LD, were employed to implement the model [104].

Seera and Lim [62], integrated a fuzzy minimum maximum neural network into classification, also into regression tree, finally to random forest in an intelligent hybrid system.

This hybrid system decreased the limitations of the fundamental models and utilized the advantages. Also, it is used the fuzzy minimum maximum neural network to incrementally learning, the classification and regression tree to predict the network outputs, and the random forest to achieve a high classification performance.

Dutta et al. [98] classified five sets of medical data which are taken from the available repository belong to California University, Irvine (UCI) using their improved firework that included an ANN model [108].

Zainuddin et al. [89] classified two data sets of epileptic seizure benchmark from California University repository by adopted a wavelet neural network. As well as, they employed an optimized harmony-search-based algorithm to select features.

Varma et al. established a technique for the purpose of handling decision trees boundaries and determining split points, which employed the Gini index in diagnosis (PID) [109].

Maddouri and Elloumi [110], introduced four distinct machine-learning techniques in order to classify the biological sequences. Luo et al. [111], discussed the recent breakthrough by using the applications of big data in health-care field. Guarracino et al. [112], trained a generalized classifier with a smaller subset of points and original data features via proposing a technique for feature selection.

The focus of most of the proposed algorithms and methods is to hybridize ANN with one or more optimization algorithms [104] or enhancing or modifying the kernel of ANN [84]; however, they do not unify the change of structural in ANN and the iterations number for metaheuristic algorithms. The effect of the selection of different iteration numbers for more than one structure of ANN is accordingly studied.

The rest of this chapter will be organized into five parts where section 2 discuss the methodology and materials which include the algorithms. In addition, it discusses the proposed work and the specification of the medical data set. Section 3 provides the experimental results while section 4 clarifies the simulation results. Finally, section 5 concludes the research and provides recommendations for future work.

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

3.2 MATERIALS AND METHODS

3.2.1 Artificial Neural Network ANN

Through transfer input data into the estimated preferred output, ANN could adopt as a classification model. This model includes three layers where the layer which receives input is named the input layer, the layer which provides the output is called the output layer while the last layer is the layer that locate between them and it is called the hidden layer (s).

The attributes from disease data sets are input to ANN in this study. These inputs will be multiplied by weights after examining them in the input layer. Relative to the connections between neurons in the hidden layer(s), the weights are prepared randomly, where the activation function uses the summation of weights, 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:

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

In order to obtain reliable estimate of classifier accuracy holdout and random sampling method used to assessing accuracy. According that the data sets in this study are separated into (40%) training, (30%) validation, and (30%) test data sets. Then, the populations are evaluated by using the trained neural network structure and the training set of weights (w1, …, wn) will be input into the metaheuristic algorithms.

3.2.2 Metaheuristic Algorithms

According to this study ANN hybridizes with three metaheuristic optimization algorithms, namely, fireworks algorithm FW, particle swarm optimization PSO and genetic algorithm GA. 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 Algorithm GA

In accordance with the fitness function, the genetic algorithm searches for an optimal solution by applying evolutionary algorithm to a set of populations, as mention in Holland [77], Mitchell [78], and Goldberg [79]. Figure 3.1 shown the algorithm functionality which are selection, crossover, and mutation operations, expressed 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- Crossover operation Performs according (crossover probability), and to methodology (single point or multi point crossover).

4- Mutation operation Performs 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 like genetic evolution that exist in nature are considered in GA.

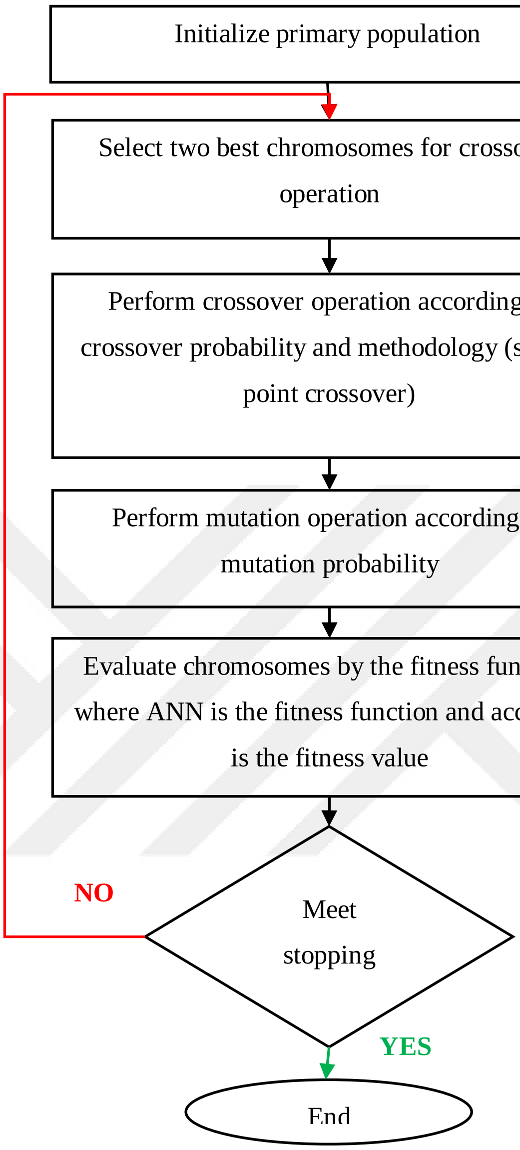


Figure 3.1: Genetic algorithm (GA)

3.2.2.2 Particle Swarm Optimization PSO

Apart from being an organized, stochastic, and self-directed algorithm Kennedy and Eberhart [95]. PSO according to Ardjani et al. [96], is a distributed population-based evolutionary technique.

Basically, the algorithm consists of set of particles, this set called swarm which can be considered as the population of solution symbols in space of search. 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.

The particles that consists of the population of solution symbols move to the suitable solution position in space of search.

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

$$\begin{aligned} \mathrm{vel}(\mathrm{it}+1) = {} & \mathrm{w} * \mathrm{vel}(\mathrm{it}) + \mathrm{s1} * \left(\mathrm{bstLcalPos} - \mathrm{pos}(\mathrm{it})\right) + \mathrm{s2} \\ & * \left(\mathrm{bstGlbalPos} - \mathrm{pos}(\mathrm{it})\right) \end{aligned} \tag{3.3}$$

$$pos(it+1) = pos(it) + vel(it+1) \tag{3.4}$$

where *vel* is the velocity, *pos* is the position of particles, *it* is the number of iterations, w is the velocity coefficient or scale that is start 1 and decreases during iterations, s1 is the fixed scale of difference between current and local positions, and s2 is the fixed scale of difference between current and global positions.

Figure 3.2 illustrates the PSO process.

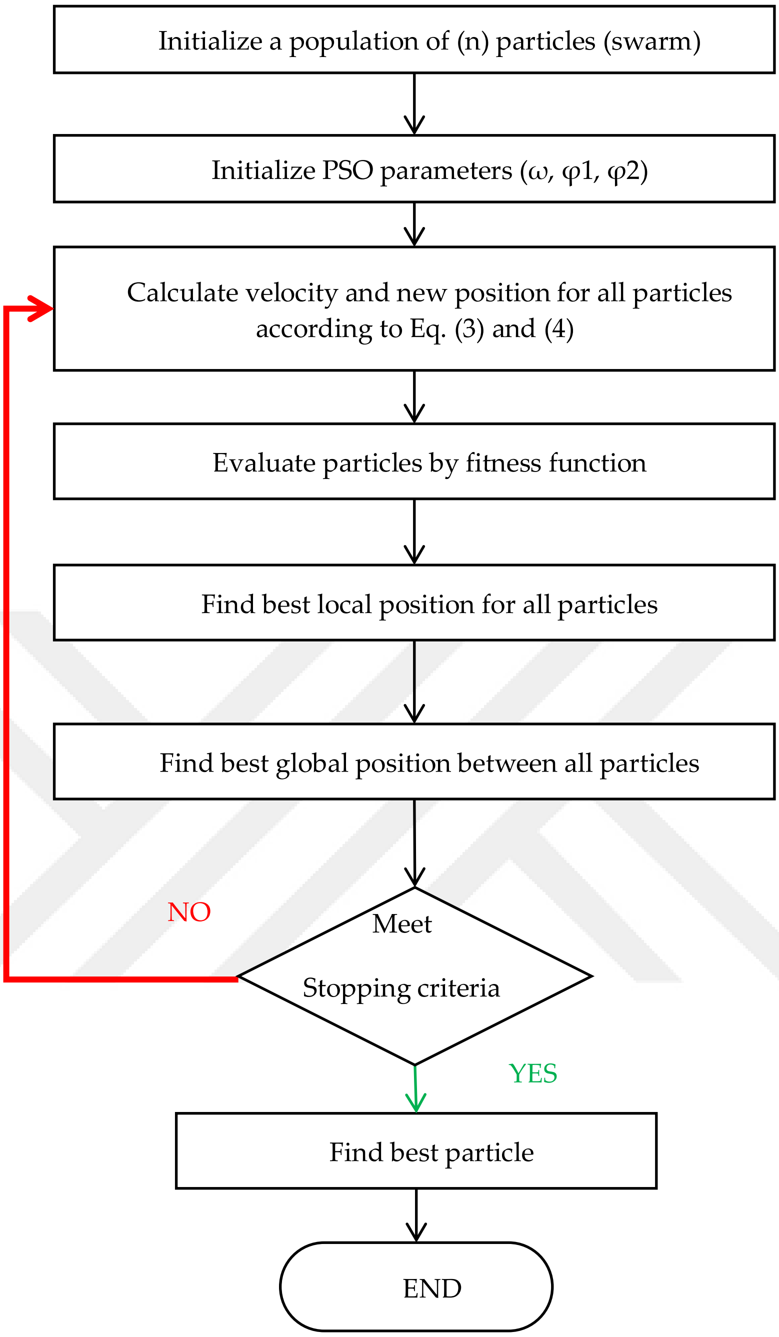


Figure 3.2: Particle swarm optimization PSO

3.2.2.3 Fireworks algorithm FW

FW 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 the fireworks as illustrated in Tan [99] and Dutta et al. [98].

Specific number of fireworks denote the initial population, the potential solutions around the population represented by the created sparks. Substantial sparks are produced in each iteration (firework), and a mutation randomly 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 sparks number for firework (Xi), the sparks maximum number represented by M, and the activation function for the firework (Xi) is f(Xi).

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

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

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

FW algorithm steps are illustrated in Figure 3.3.

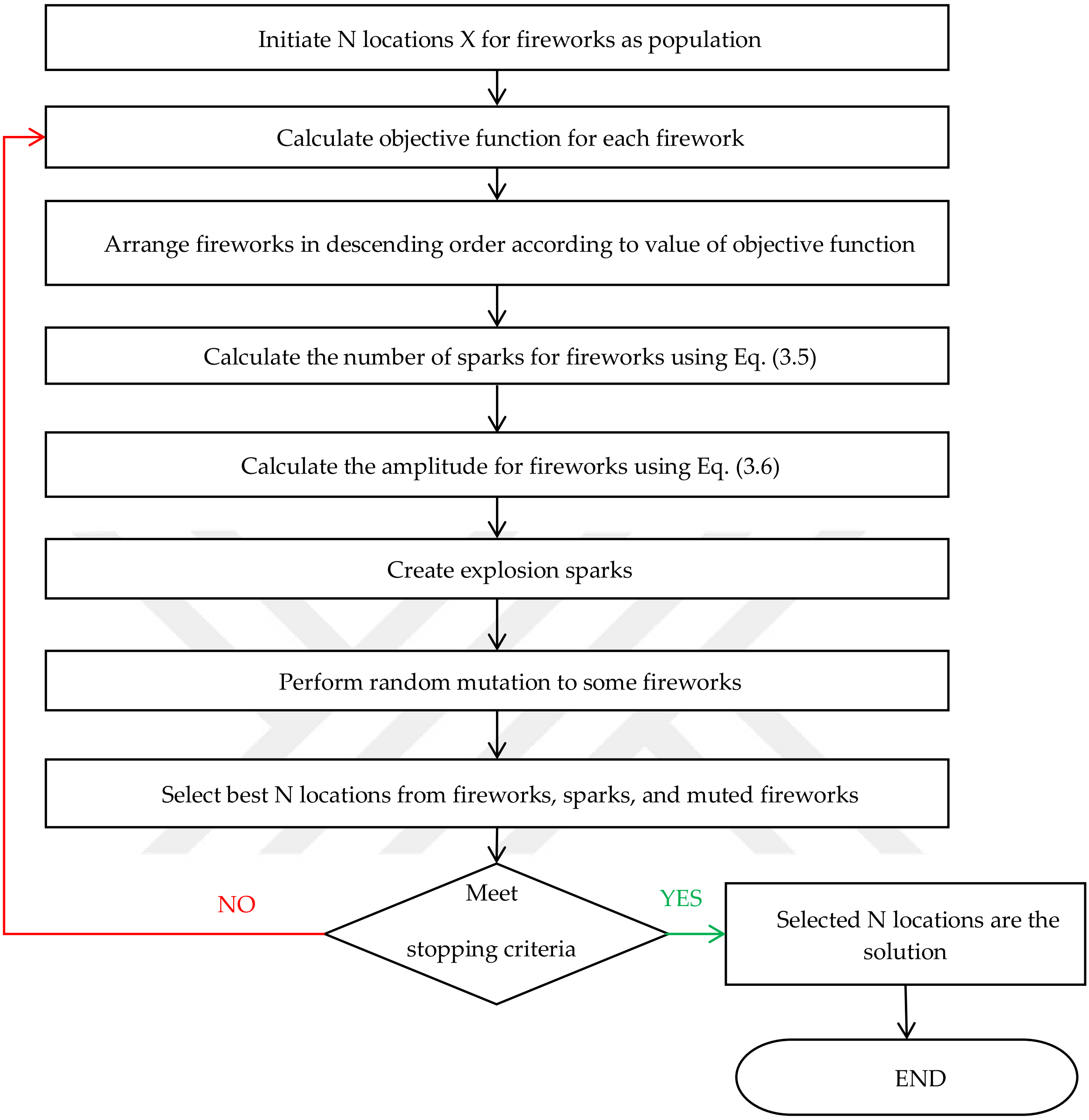


Figure 3.3: Fireworks approach flowchart (FW)

3.2.3 Proposed Work

This study hybridizes three metaheuristic approaches, namely, PSO, GA and FW with ANN. The best weights set of ANN is determined for constructing these methods to improve the accuracy and enhance performance. The general structure of this study is presented in Figure 3.4.

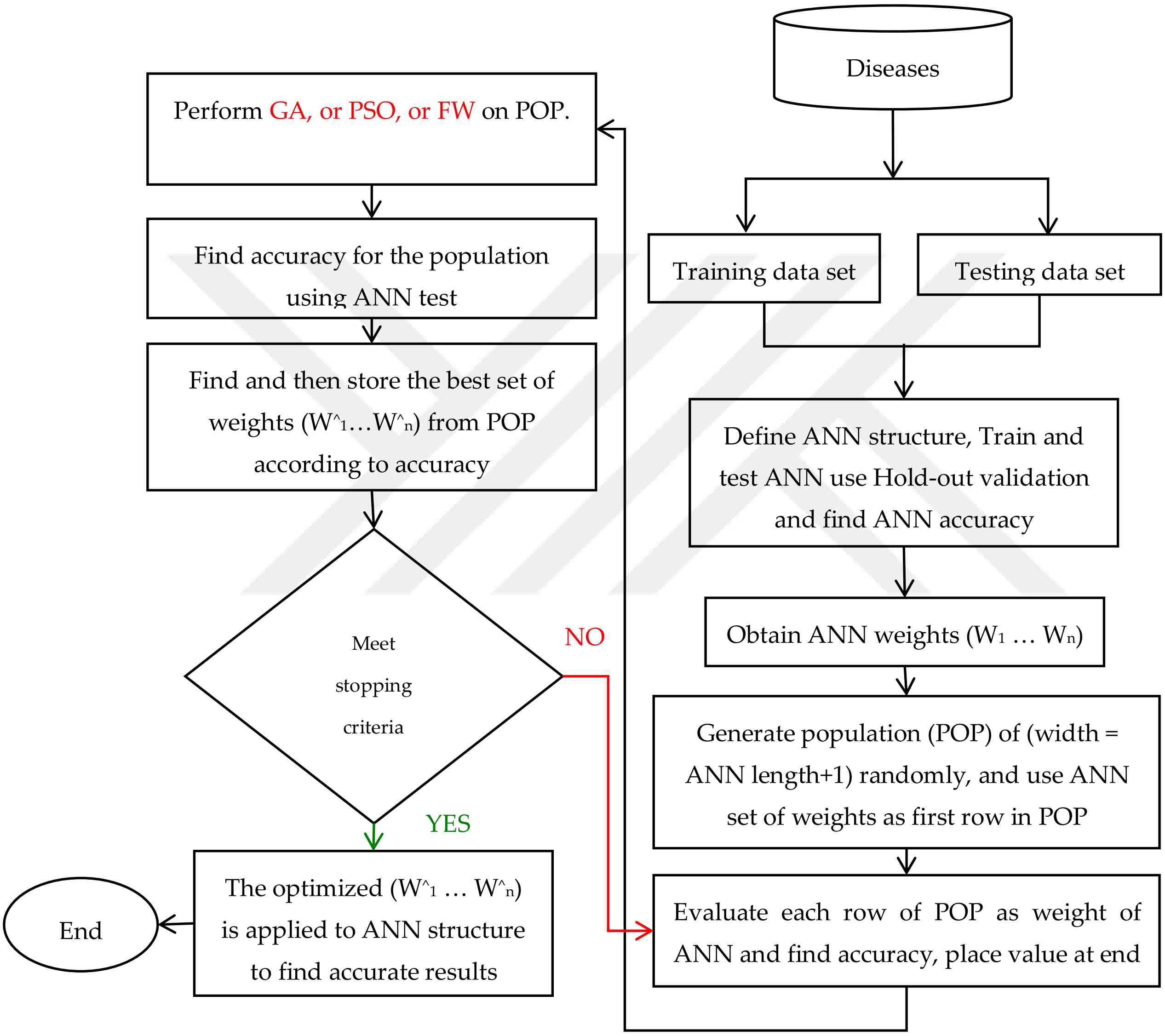


Figure 3.4: The proposal methodology: ANN+GA, ANN+PSO, ANN+FW.

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

3.2.4 Datasets

Optimization algorithms are hybridized with ANN in order to improve the following disease data sets classification [108].

(1) PID

This data set comprises pregnant female's patients with at least 21 years old and of the heritage from Indian Pima, as record in the National Institute of the Diabetes and the Digestive and Kidney Diseases.

Record numbers: 768

Tests and diagnose (attributes number): 9 (comprising class)

(2) WDBC

In this data set a description of the cell nuclei features that exist in the image. The University of Wisconsin employs a digitized image of a fine needle aspirate of a breast mass for the purpose of computing the characteristics.

Patients numbers: 569

Tests and diagnose (number of attributes): 32

(3) LD

Seven attributes are included in this data set. The results of blood test associated with the disorder of liver due to consumption of alcohol are indicated in the five attributes. The sixth attribute is constituted by the number of drinks in each day. While the seventh attribute presents the class and conditions of patient including whether the patient has disorder or not.

Patients numbers: 345

Tests and diagnose (attributes number consisting class attribute): 7

(4) HSS

A study conducted in Billings Hospital which belong to University of Chicago on the survival status of patients who experienced surgery related breast cancer disease between 1958 and 1970. Results were gathered in this data set.

Patients numbers: 306

Tests and diagnose (attributes number consisting class attribute): 4

(5) PD

Max Little from the University of Oxford formed this data set Through cooperation with the National Center for Voice and Speech that locates in Denver, Colorado. Twenty-three sound measures consider as attributes are contained in this data set. As well as, they provided a total of 197 patients that correspond to sound records of 31 individuals, 23 of whom have Parkinson’s disease.

(6) RNA-Seq gene expression cancer

RNA-Seq extracted randomly gene expression of patients with many types of tumor such as colon tumor COAD, prostate tumor PRAD, breast tumor BRCA , lung tumor LUAD and kidney tumor KIRC.

This data set is part of TCGA The Cancer Genome Atlas Pan-cancer Analysis Project by Yuan et al. [113]. The RNA-Seq data set has been declared through the literature especially in the scientific paper of Cestarelli et al. [114].

The Illumina HiSeq platform (DNA-to-Data solutions) measures the RNA-Seq gene expression levels to represent gene expressions as numerical sequences within the data set.

Patients numbers: 801

Attributes number: 20531

Classes number: 5

The data sets specification is shown in Table 3.1.

Table 3.1: Medical datasets specification

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
Gene expression cancer (RNA-Seq)	801	20531	5 cancer types: BRCA, KIRC, PRAD, LUAD, COAD	

3.2.5 The Strategy for solving the problem

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

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

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

The following is algorithm parameters after performing several tests:

(1) Setting GA parameter

Population	10
Probability of crossover	0.7
Probability of mutation	0.1

(2) Setting PSO parameter

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) Setting FW parameters

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 OF THE STUDY

3.3.1 First Experiment

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

Hybridizing PSO, GA, and FW with ANN for the five diseases data sets in this experiment design is as follows:

- 500 iterations with two hidden layers for ANN (Table 3.2).
- 1000 iterations with two hidden layers for ANN (Table 3.3)
- Record enhancement.
- 500 iterations with three hidden layers for ANN (Table 3.4)
- 1000 iterations with three hidden layers for ANN (Table 3.5)
- Record enhancement.

3.3.2 Results of Experiment 1

Empirical implementation conducted according experiment 1, results registered in the following tables, table 3.2 and table 3.3 provide the results of implementing 2 hidden layer ANN with the three proposed algorithms. Table 3.4 and table 3.5 provide the results of implementing 3 hidden layer ANN with the three proposed algorithms.

Table 3.2: Five medical benchmarks classification by hybridize two ANN hidden-layers with 500 iteration metaheuristic algorithms

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 + FW	79.56	**98.07**	73.07	78.76	89.23

Table 3.3: Five medical benchmarks classification by hybridize two ANN hidden-layers with 1000 iteration metaheuristic algorithms

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 + FW	**79.95**	98.07	73.33	78.43	91.28

The following improvements for ANN are indicate in table 3.2 and table 3.3:

- Metaheuristic GA: enhancement in 4 out of 5 benchmarks
- Metaheuristic PSO: enhancement in 4 out of 5 benchmarks
- Metaheuristic FW: enhancement in 3 out of 5 benchmarks

Table 3.4: Five medical benchmarks classification by hybridize three ANN hidden-layers with 500 iteration metaheuristic algorithms

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 + FW	78.39	**98.42**	70.14	77.78	89.28

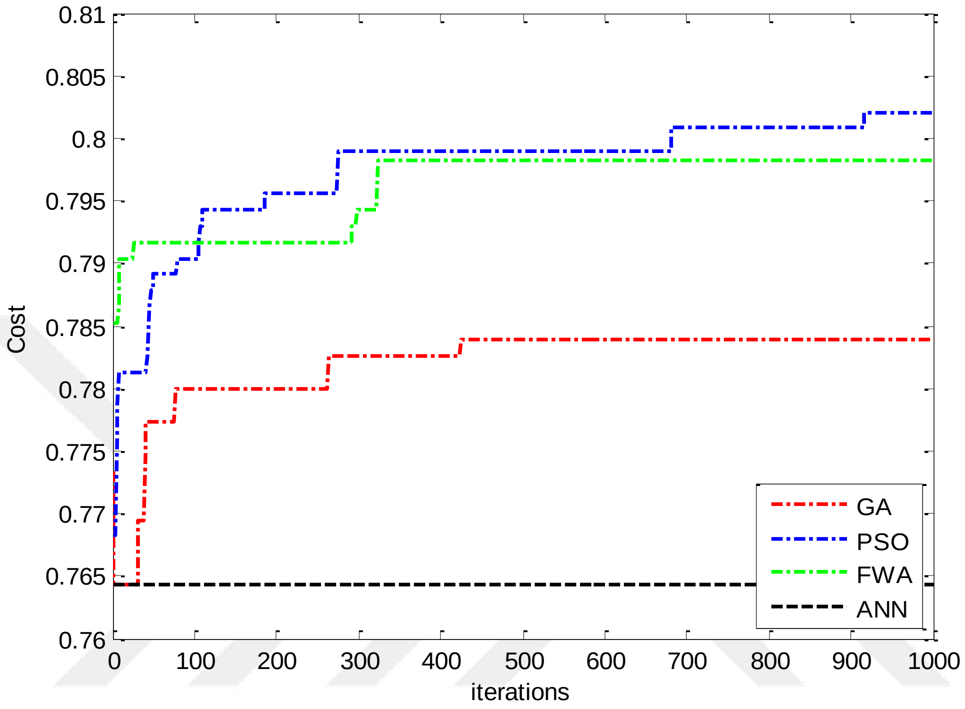
Table 3.5: Five medical benchmarks classification by hybridize three ANN hidden-layers with 1000 iteration metaheuristic algorithms

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 + FW	79.82	**98.42**	73.91	78.10	90.77

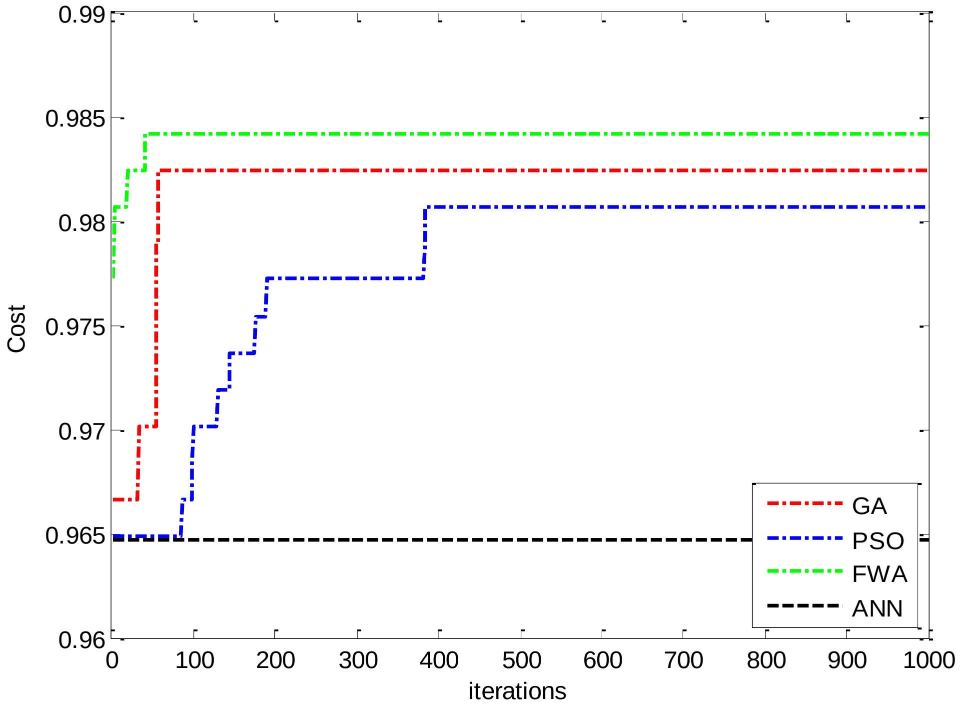
The following improvements for ANN are indicate in table 3.4 and table 3.5:

- Metaheuristic GA: enhancement in 3 out of 5 benchmarks
- Metaheuristic PSO: enhancement in 4 out of 5 benchmarks
- Metaheuristic FW: enhancement in 4 out of 5 benchmarks.

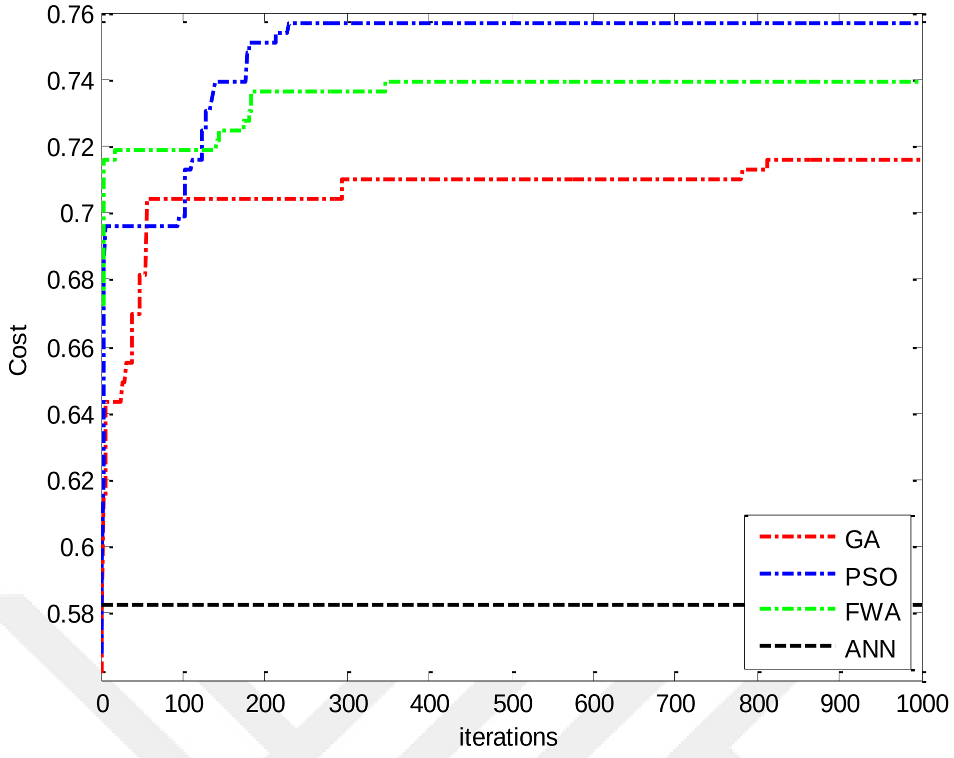
Figure 3.5, illustrate the performance of separately proposed algorithms using 1000 iterations in enhancing ANN with three hidden layers. The best performance in most cases is given by the PSO with 1000 iterations and three-hidden-layer ANN. Therefore, the most promising algorithm relative to enhancement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



(1) PID

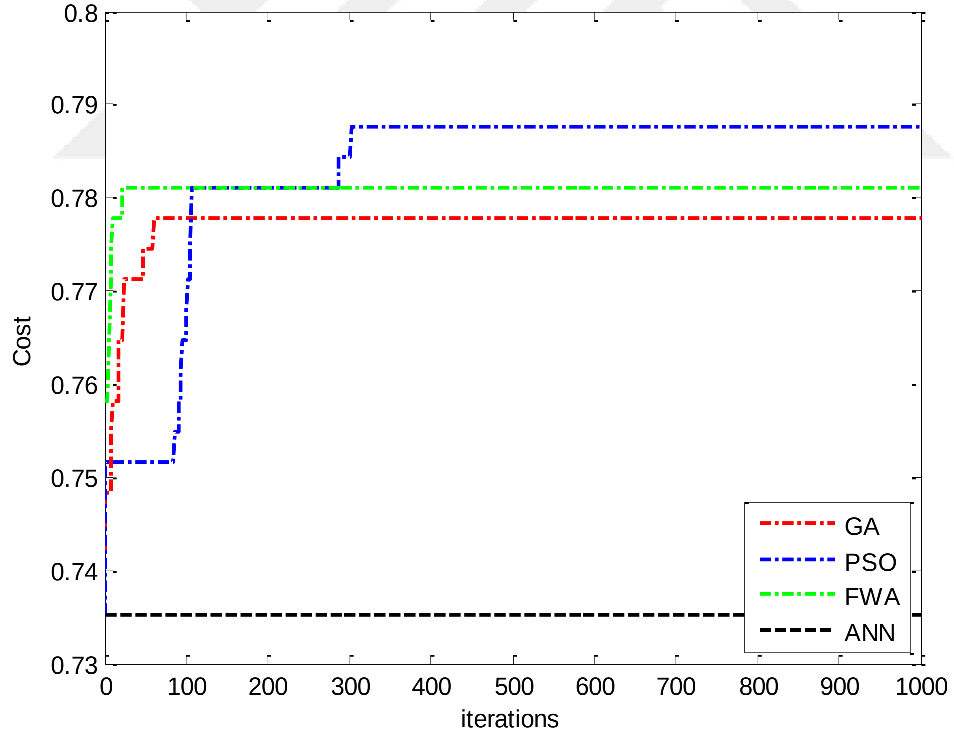


(2) WDBC



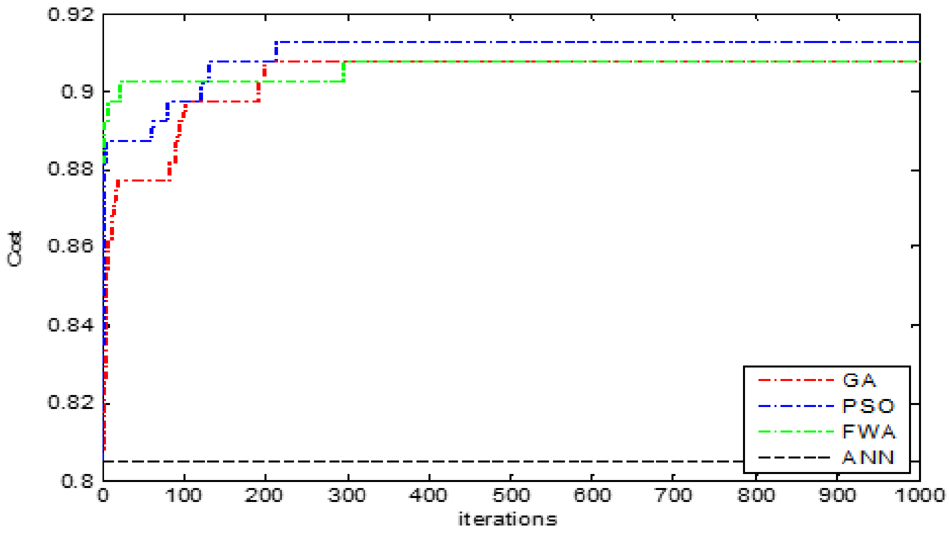
Cost
0.76
0.74
0.72
0.7
0.68
0.66
0.64
0.62
0.6
0.58
0
100
200
300
400
500
600
700
800
900
1000
iterations
GA
PSO
FWA
ANN

(4) LD



Cost
0.8
0.79
0.78
0.77
0.76
0.75
0.74
0.73
0
100
200
300
400
500
600
700
800
900
1000
iterations
GA
PSO
FWA
ANN

(5) HSS



(6) PD

Figure 3.5: Illustration of accuracy comparing in 1000 iteration for 3 optimization algorithms, GA, PSO, and FW hybridize with 3 hidden layer ANN tested on the medical data sets: (1) PID; (2) WDBC; (3) LD; (4) HSS; (5) PD

3.3.3 Experiment 2

PSO achieves the superlative accuracy results in the first experiment on the mention benchmark disease datasets, except WDBC. However, PSO comprises the most attributes among the 30 features of WDBC. In experiment 2, only PSO is used to enhance ANN, and parameters are tuned with different ANN architectures.

The initialization condition of PSO is changed in this experiment. The swarms entirely initialize round the weights set acquired from the training phase of initial ANN. Then, the PSO searches locally to increase intensification. The following strategies are implemented for hybridization to survey search space and ensure exploration (diversification).

- Firstly, 1000 iterations of PSO and 100 swarms for three ANN hidden layers
- Secondly, 1000 iterations of PSO and 100 swarms for four hidden layers
- At last, 1000 iterations of PSO and 200 swarms for four ANN hidden layers

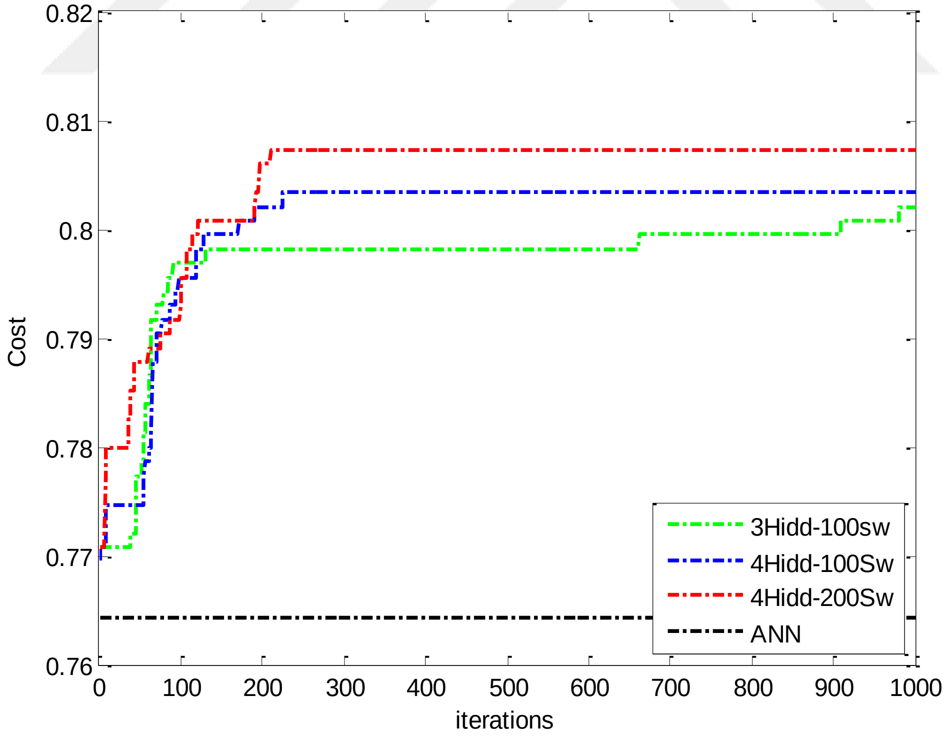
3.3.4 Experiment 2 results

Table 3.6, presents the results of experiment 2. PSO performances with four hidden layers by the use of 1000 iterations and 200 swarms of PSO population implemented on the medical benchmarks are demonstrated.

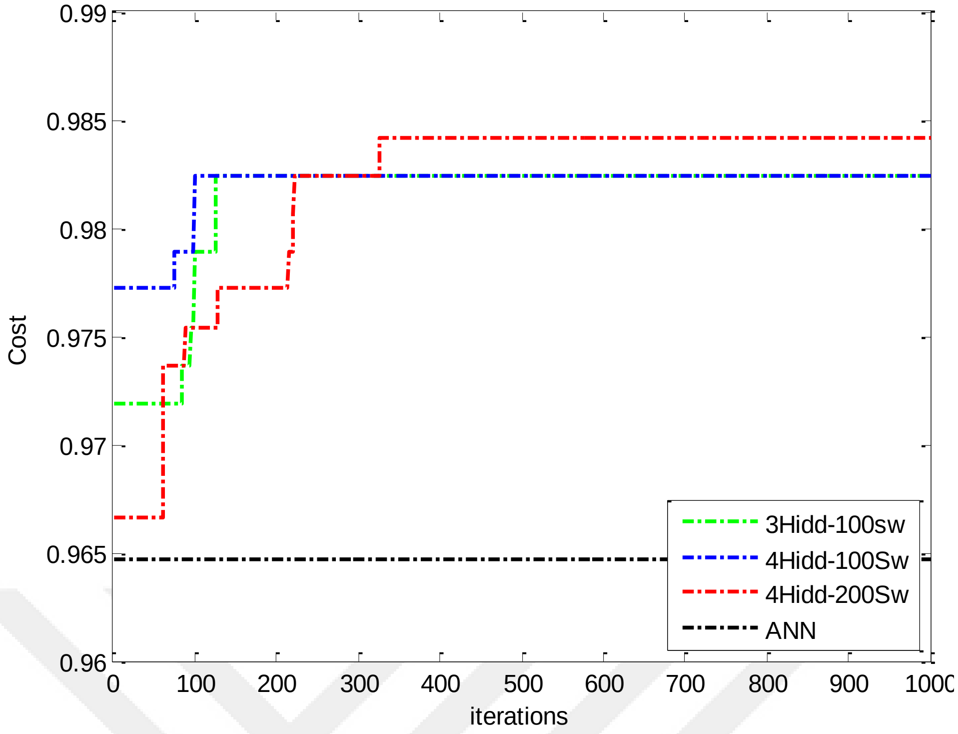
Table 3.6: Results of experiment 2, PSO algorithm for 1000 iterations to enhance 3 and 4 hidden-layer ANN

2 Hidden_ 500 Iteration		PID	WDBC	LD	HSS	PD
Classical ANN		74.35	95.39	57.10	72.88	75.38
Hybridized	ANN + PSO(3,1000), 100 swarm	80.21	98.42	74.49	78.43	90.26
	ANN + PSO(4,1000), 100 swarm	79.56	97.01	76.52	78.43	91.28
	ANN + PSO(4,1000), 200 swarm	**80.34**	**98.42**	**76.52**	**79.41**	**92.31**

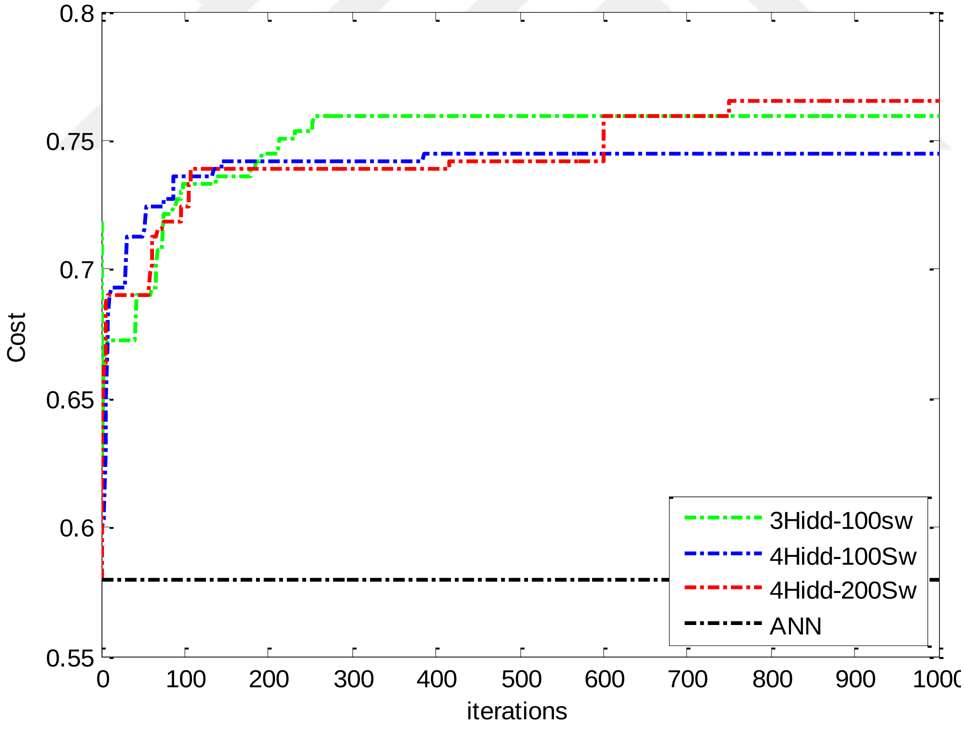
The experimental results of the five medical datasets classified by 3 hidden layer ANN enhanced with the three population metaheuristic algorithms named PSO, GA, and FW are illustrate in the below figure 3.6 (1, 2, 3, 4, and 5).



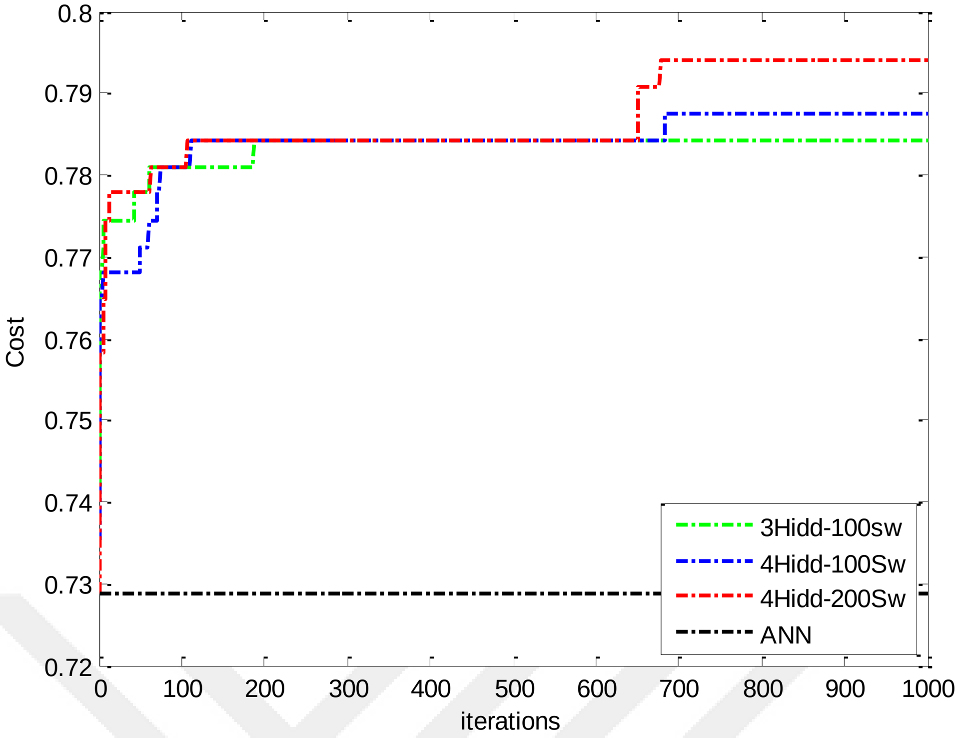
(1) PID


Cost
0.99
0.985
0.98
0.975
0.97
0.965
0.96
0
100
200
300
400
500
600
700
800
900
1000
iterations
3Hidd-100sw
4Hidd-100Sw
4Hidd-200Sw
ANN

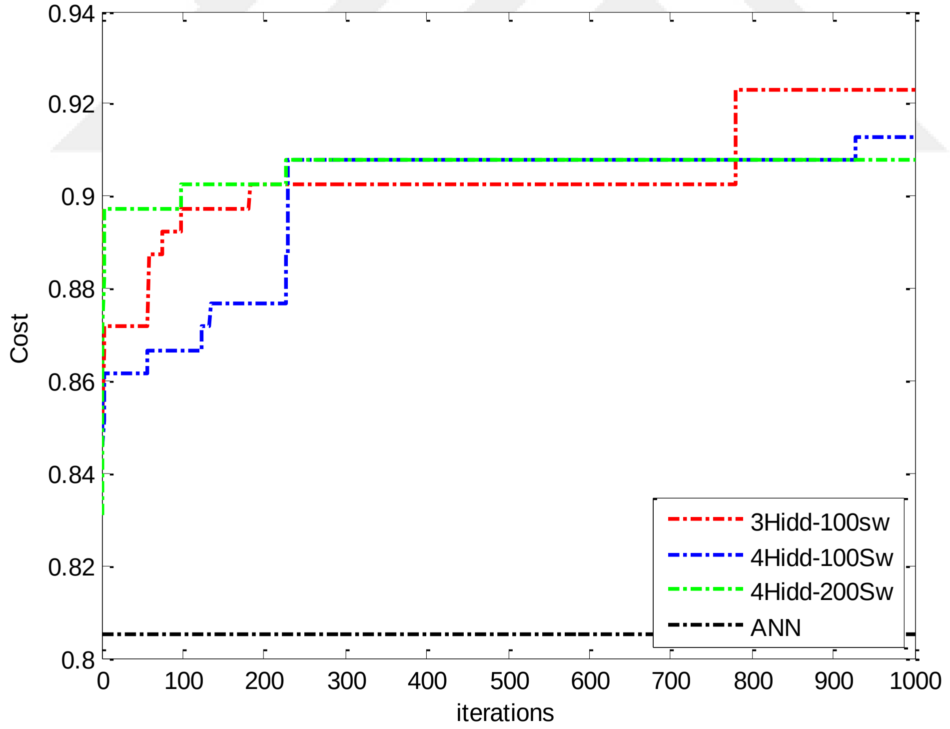
(2) WDBC


Cost
0.8
0.75
0.7
0.65
0.6
0.55
0
100
200
300
400
500
600
700
800
900
1000
iterations
3Hidd-100sw
4Hidd-100Sw
4Hidd-200Sw
ANN

(3) LD



(4) HSS



(5) PD

Figure 3.6: PSO performances using 1000 repeated iterations and 200 population swarms with four hidden layers, tested on the medical data sets: (1) PID; (2) WDBC; (3) LD; (4) HSS; (5) PD

The two algorithms means are compared by a statistical hypothesis test called the t-test. The differences of the two algorithms is determine by apply correlation and regression to them. The t-test statistics is changed into a conditional probability that is named as p-value. P-value answers the query, “If the null hypothesis is true, then what is the probability to observe the current data or data that are more extreme?”

When the null hypothesis is true, the observed data are not similar. Therefore, a small p-value presents indication against the null hypothesis Afifi and Azen [115].

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

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

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

where df pertains freedom degree. While β denotes the beta function.

On the basis of defining the variances between ANN and optimized ANNPSO, the statistical assessment on the 5 diseases benchmarks is significant as explained in Table 3.7.

Table 3.7: ANN and ANN + PSO Statistical test

	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

3.3.5 RNA-Seq big data test using hybridized Algorithms

This study uses big medical data sets, the dataset represents patients records with more than 20,000 gene test for each person, namely the RNA-Seq Cancer Data Set as in Cestarelli et al.

[114]; the Cancer RNA-Seq Data Set randomly extracts gene expression of infected people with the types of tumor of PRAD, LUAD KIRC, BRCA and COAD.

The three metaheuristic algorithms are hybridized with a four-hidden-layer ANN and implemented for classification on the RNA-Seq. The three hybridized algorithms show improvements, but GA is superior to the other optimization algorithms as illustrate in Table 3.8 .

Table 3.8: RNA-Seq accuracy results from ANN and the three algorithms

	ANN	ANN + PSO	ANN + GA	ANN + FW
RNA-Seq	93.68	95.03	**98.75**	96.31

Results of the statistical tests shown in Table 3.9 showing the significance of FW and GA against the insignificance of PSO statistically with ANN.

Table 3.9: ANN and the three algorithms statistical test of RNA-Seq

	ANN + GA	ANN + FW	ANN + PSO
Mean difference	0.004328	0.033836	−0.007176
T score	2.9800	7.6776	0.4819
standard error of difference	0.001	0.004	0.015
Difference verdict	Significant	Significant	Not Significant

3.4 DISCUSSION

An effective factor in ANN learning is the hidden layers number. If the number of these layers are increased, ANN complexity will be increased accordingly and overfitting in ANN learning. Nevertheless, when this number is increased, the model can obtain precise classification results. Metaheuristic algorithms are employed to generalize the ANN and in order to solve the problem of overfitting and enhance the classification.

Metaheuristic approaches, which are search approaches, and consider an improve mathematical models to yield well 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 in the interior of a single iteration of the algorithm, and mathematical models (the ANN model in this study) can search for improved solutions within the search space by increasing the iterations number.

The following observations are acquired from experiment 1:

The following benchmarks are enhanced by increasing the iterations number in two hidden layers ANN from 500 to 1000 as indicate in table 3.2 and table 3.3:

1. In GA: 4 out of 5
2. In PSO: 4 out of 5
3. And in FW: 3 datasets out of 5 datasets enhanced.

- The following benchmarks are enhanced by increasing the iterations number in two hidden layers ANN from 500 to 1000 as indicate in table 3.4 and table 3.5:

(1) GA: 3 out of 5
(2) PSO: 4 out of 5
(3) FW: 4 out of 5

According to experiment 1, the hybridization ANNPSO algorithm is considered the supreme appropriate methodology for this study.

The finest approach from experiment 1 which is PSO is used and its parameters are fine tuned in experiment 2. We begin from the initialization of population. The population initialization of the three algorithms in experiment 1 is divided into two processes. First, initialization is conducted with the ANN weights set ($w_1…, w_n$) and neighbors around. Second, random initialization is performed. In the second experiment, the whole primary populations are extracted from the ANN set of weights and neighbors.

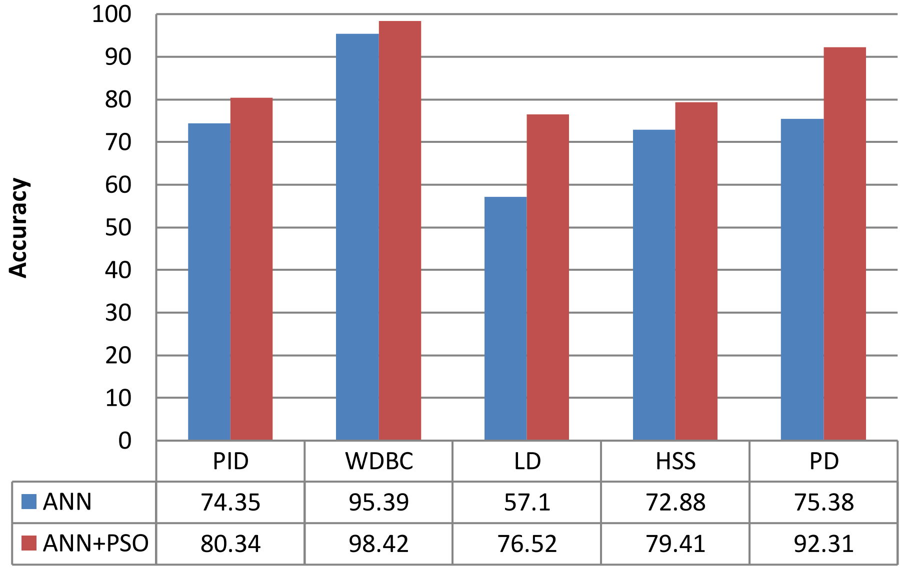
Better solutions in the hyperspace can be determined via population density. Therefore, 100 population and 200 population swarms are tested, as explained in section 3.3.3 the second experiment, this illustrated in table 3.6. Experiment 2 provides the following empirical improvement results between ANN and PSO comparing with experiment 1:

- First, 3 hidden layers, 100 swarms: Two out of five medical data sets.
- Second, 4 hidden layers, 100 swarms: Two out of five medical data sets.
- Finally, 4 hidden layers, 200 swarms: All medical data sets enhanced.

PSO is determined to be the best algorithm in this study as stated by the aforementioned accuracy results. The results of this algorithm are comparable to methods in literature, as indicated in table 3.10. the enhancement in the PSO algorithm with ANN on the five medical benchmarks depicted in figure 3.7.

Table 3.10: Compare proposal with the state-of-the-art results

Pima Indian Diabetes (PID)	77.60% Au et al. [116]	77.60%
	Luukka [117]	75.29%
	Örkcücet al. [104]	77.60%
	proposed	80.34%
Wisconsin Breast Cancer (WDBC)	Örkcü et al. [104]	94.00%
	Luukka [117]	98.19%
	Seara & Lim [62]	98.84%
	proposed	98.42%
Liver Disorder (LD)	Luuka [117]	70.25%
	Lee and Mangasarian [118]	74.86%
	proposed	76.52%
Haberman Surgery Survival (HSS)	Pham [10]	72.70%
	Yang et al. [119]	51.96%
	proposed	79.41%
Parkinson’s (PD)	Luuka [117]	85.03%
	Rustempasic & Can [120]	81.34%
	Shrivastava [100]	93.60%
	proposed	92.31%



	PID	WDBC	LD	HSS	PD
🟦 ANN	74.35	95.39	57.1	72.88	75.38
🟥 ANN+PSO	80.34	98.42	76.52	79.41	92.31

Figure 3.7: Experiment 2 enhancement

There is significant difference between the accuracies of ANN and ANNPSO on the five medical data sets are determined in the statistical test.

The aforementioned results do not necessarily indicate that the hybridized PSOANN consider as common method for all medical problems and this belief is not correct. The no–free-lunch theorem NFLT (Ho and Pepyne [36]; Wolpert and Macready [35]) states that no greater enhancement algorithm is existed for the entire problems. This conclusion is acquired from implementing the same optimization approaches on the RNA-Seq data set. PSO, which is competitive on the first five data sets, fails with the large data in the RNA-Seq.

The effects of population-metaheuristic iterations number on varying ANN structures and tune-metaheuristic parameters are evaluated in this study. The ability of the population methods to classify the large biomedical data set is tested. The experiment with the no-free-lunch theorem is then conducted. The conclusions of this study can be extended to diverse fields and applications in future work. For example, the effect of selecting diverse ANN core functions on hybridization, or when applying various forms of metaheuristic algorithms like trajectory algorithms, can be studied. Metaheuristic and machine learning can be hybridized to improve metaheuristic mathematical models for trade-off between exploration and exploitation.

3.5 CONCLUSIONS

ANN is hybridized with three optimization approaches through this study. They are GA, PSO, and FW. The three algorithms are adopted for ANN hybridization with diverse layer structures, and the parameters of metaheuristic algorithms are tuned. Unlike most of other approaches in the literature, PSO is competitive on the five utilized medical benchmarks

Table 3.10. Whereas, classification experiment of the RNA-Seq is a hands-on implementation of the no-free-lunch theorem. This theorem states that no universal common optimization algorithm for the entire optimization problems which existed. Our approach can be adopted in future studies to solve various medical problems.

4. EFFECT OF DIVERSIFICATION AND INTENSIFICATION TRADE-OFF IN ANEMIA MEDICAL DATA CLASSIFICATION

4.1 INTRODUCTION

In this chapter, the (ANN) was hybridized with a hybridization between differential evolution algorithm (DE) and simulated annealing (SA) to improve anemia disease classification. DE and SA are metaheuristic algorithms; DE is considered as an evolutionary population-based solution algorithm, whereas SA is a trajectory method operating on a single solution. Yang et al. [121], Blum and Roli [13], and Talbi [76] stated that the balance between the intensification and diversification is considered the major role of metaheuristic algorithms.

Diversification (exploration) performs well with evolutionary metaheuristic algorithms, whereas trajectory algorithms are considered the best in local search (exploitation). In this work, DE and SA were hybridized to enhance the ANN for anemia medical dataset classification. The proposed methodology begins with choosing the best ANN structure (best of hidden layers number and best neurons number in each layer) and then enhancing the classification process by hybridizing DE and SA to enhance the accuracy of classification. The proposed methodology registered a significant enhancement in anemia classification. The anemia data is a real dataset gathered from Iraqi blood laboratories to detect anemia diseases. In addition, the proposal was applied into two benchmarks from the UCI repository, Pima Indian diabetes PID data, and liver disorder LD diseases. The proposal method registered remarkable results.

Many different data-mining algorithms in literature are used to classify several types of diseases, such as anemia disease, into specific types on the basis of the data-mining algorithms by Elshami and Alhalees [122]. A person with anemia is probably unaware of the problem because symptoms may not appear. Millions of people may suffer from anemia and their health exposed to risk. Therefore, the disease is significant; several studies carried out in this domain are mentioned in the study of Yilmaz et al. [123]. Sanap et al. [124] developed a system by using the classification technique of WEKA, they are decision tree C4.5 algorithm and SVM. The scholars performed a number of experiments that uses these algorithms. The anemia classification applies a decision tree that provides clear results depending on complete blood count (CBC) reports. Amin et al. [125], have compared between naive Bayes, J48 classifier, and neural network classification algorithms through WEKA and studied hematological data to specify the most appropriate algorithm.

Yang et al. [121] stated that there is no rule can promise the balance between exploration (diversification) and explanation (intensification) in metaheuristic algorithms for all problems. However, different studies attempt to achieve a balance between diversification and intensification. Fagan and Vuuren [126] declared six general views of diversification and intensification terminology from literature reviews. Another study by Makas and Yumusak [127] combined the artificial bee colony (ABC) with migrating birds optimization (MBO) for the purpose of attaining balance between exploration and explanation through the exploration property of ABC and MBO explanation property under a strategy of the sequential execution.

4.2 METHODS AND MATERIALS

4.2.1 Artificial neural network ANN

Through transfer input data into the estimated preferred output, ANN could adopt as a classification model. This model includes three layers where the layer which receives input is named the input layer, the layer which provides the output is called the output layer while the last layer is the layer that locate between them and it is called the hidden layer (s).

The attributes from disease data sets are input to ANN in this study. These inputs will be multiplied by weights after examining them in the input layer. Relative to the connections between neurons in the hidden layer(s), the weights are prepared randomly, where the activation function uses the summation of weights, as indicated in Eq. (4.1) and Eq. (4.2)

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

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

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

In order to obtain reliable estimate of classifier accuracy holdout and random sampling method used to assessing accuracy. According that the data sets in this study are separated into (40%) training, (30%) validation, and (30%) test data sets. Populations evaluation done by the structure of the trained neural network, and the training weights set (w1, …, wn) are input into the metaheuristic algorithms.

4.2.2 Differential evolution DE

The DE initializes the random population of d-dimensional vectors. The representation of the solution applied to the DE is the same as that applied in the GA.

The DE combines several solutions with the candidate solution to create a new solution, as stated by Storn and Price [90]. Three main operations were considered for population solution evolution through repeated cycle; these operations were mutation, crossover, and selection. Despite the similarity in naming GA operations, the operations were not exactly the same.

The process of generation in every iteration followed those of Storn and Price [90]; Kachitvichyanukul [128]:

Three vectors were selected randomly from the population (not the target vector) and combined to generate mutant vector V as the first step. The combination process was performed in accordance with Eq. (4.3).

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

where X_1, X_2, and X_3 are the randomly nominated vectors from the population, and f refers to the constant factor that control the differential variation amplification ($X_2 - X_3$) and considered as the main parameter of DE.

The second step is the crossover between mutant and target vectors. The crossover method in DE was used for either binomial crossovers in Eq. (4.4) to produce trial vector according specified crossover probability CR.

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

where the trial vector is U, the mutant vector is V, and X is the target vector. CR is the crossover rate, whereas i: 1 ... D, where D is considered as the number of dimensionalities. Figure 4.1 show crossover process for 8-dimension parameters.

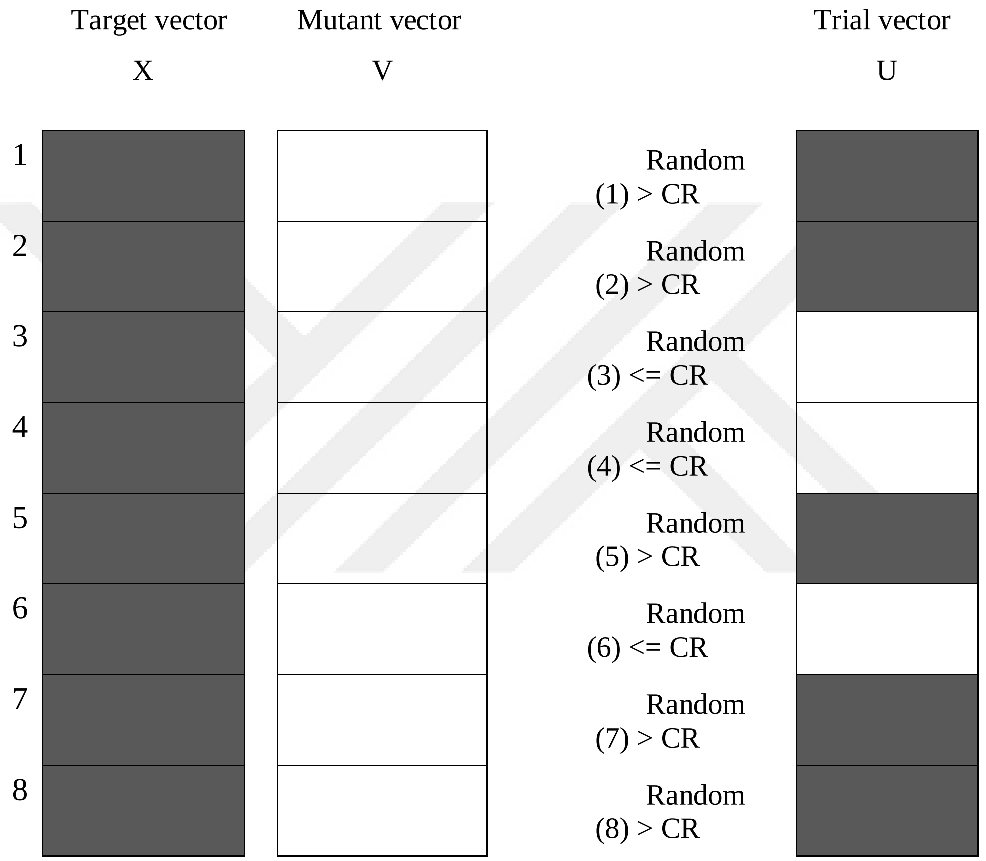


Figure 4.1: Crossover process for 8 parameters

The third step involves the selection operation where the best vector is chosen between trial vector and the target vector depending on fitness. The best fitness is targeted in the next generation. Figure 4.2, illustrate operation steps of differential evolution algorithm as flowchart.

In this study, each individual represents a set of weights for the ANN model, and fitness is reflected by the ANN accuracy.

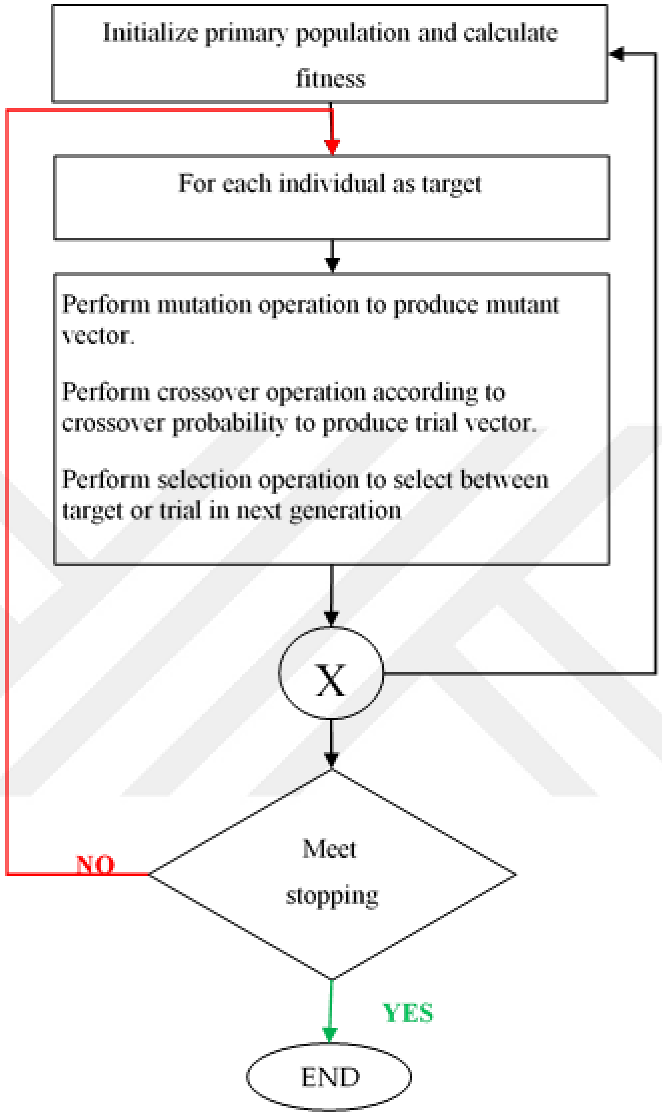


Figure 4.2: DE flowchart

4.2.3 Simulated Annealing SA

SA functionality is using one representative or solution that moves over the design or space of search in a piecewise style, as described by Kirkpatrick et al. [129]. Basically, SA is built on a previous method called "Metropolis algorithm" by Metropolis et al. [72]; Blum and Roli [13], In the latter algorithm, some jobs are accepted that do not lower the range when they attend to let the agent exploring further the possible solutions space. Such "bad" trades are permitted under the standard in Eq.(4.5):

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

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

In this study, SA's steps assist to jump local minima solutions and enforce the algorithm to explore global optimum solution as in figure 4.3. The SA algorithm adopted the steps illustrated in the flowchart in figure 4.4, as shown in the flowchart the search will depend on the value of $\boldsymbol{e}^{(-\frac{newsol-sol}{Temp})}$, which decreases over time to provide diversification for the solution through algorithm loop iteration.

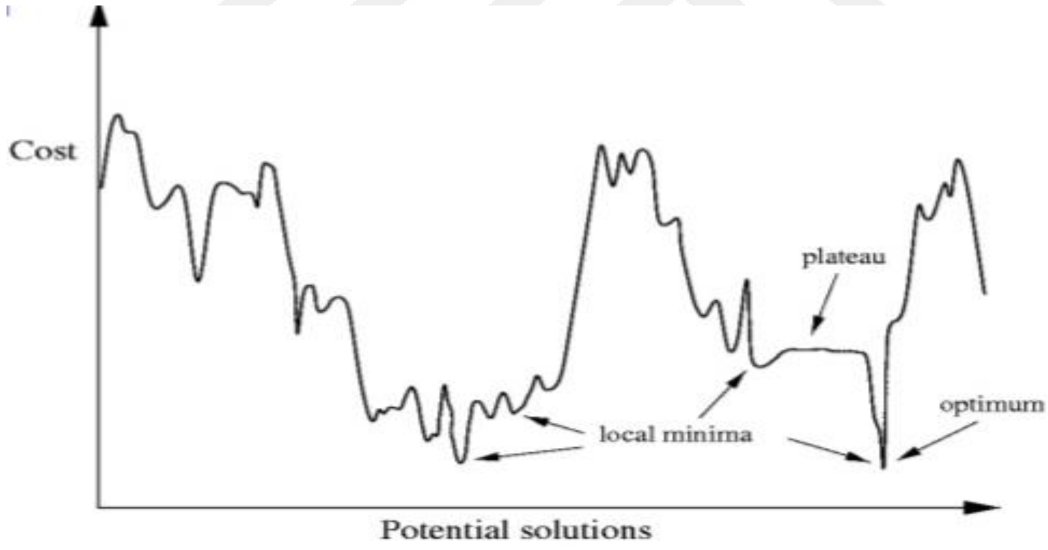


Figure 4.3: SA strategy to jump local minima

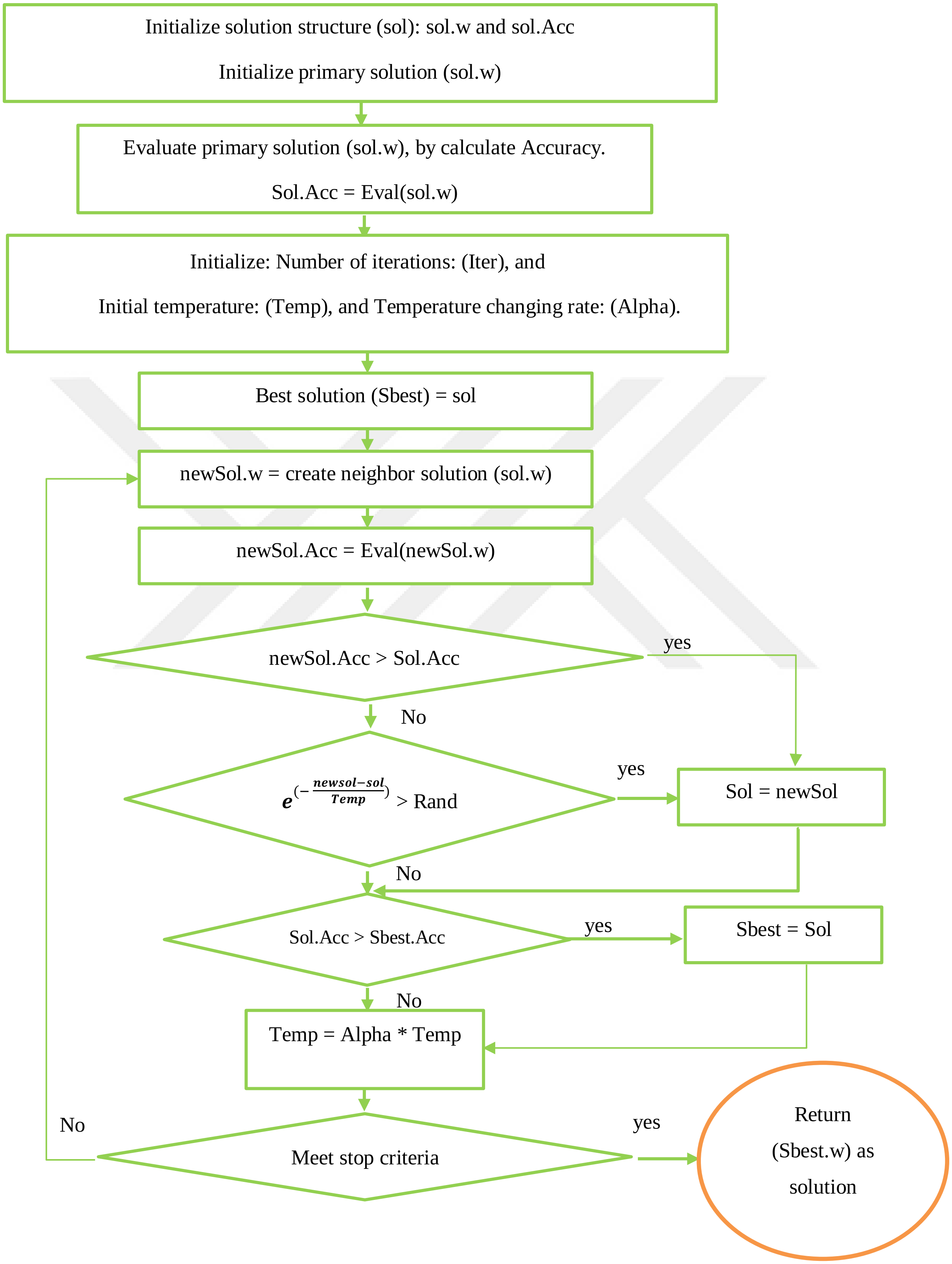


Figure 4.4: Simulated annealing algorithm

4.2.4 Proposed Work

This study presents the details of hybridization between differential evolution (DE) and SA to enhance the ANN and proposes the ANN_DESA method to balance exploration (diversification) with exploitation (intensification). The proposed methodology begins with choosing the best ANN structure, best of hidden layers number, and best number of neurons in each layer. The selection is performed by testing every ANN structure on the dataset with boundaries of 30 hidden layers and 10 neurons to find the best structure with the best accuracy. Weights of the best structure (W1 … Wn) were introduced to hybridize DESA, where DE adopted an exploration role in the method in terms of diversification solutions, whereas SA maintained a diversification role for exploitation in the promising solution search space as details in Figure 4.5. Resulting weights (W^1 … W^n) were tested on dataset selection for enhanced accuracy. **Figure 4.6** illustrates the proposed study.

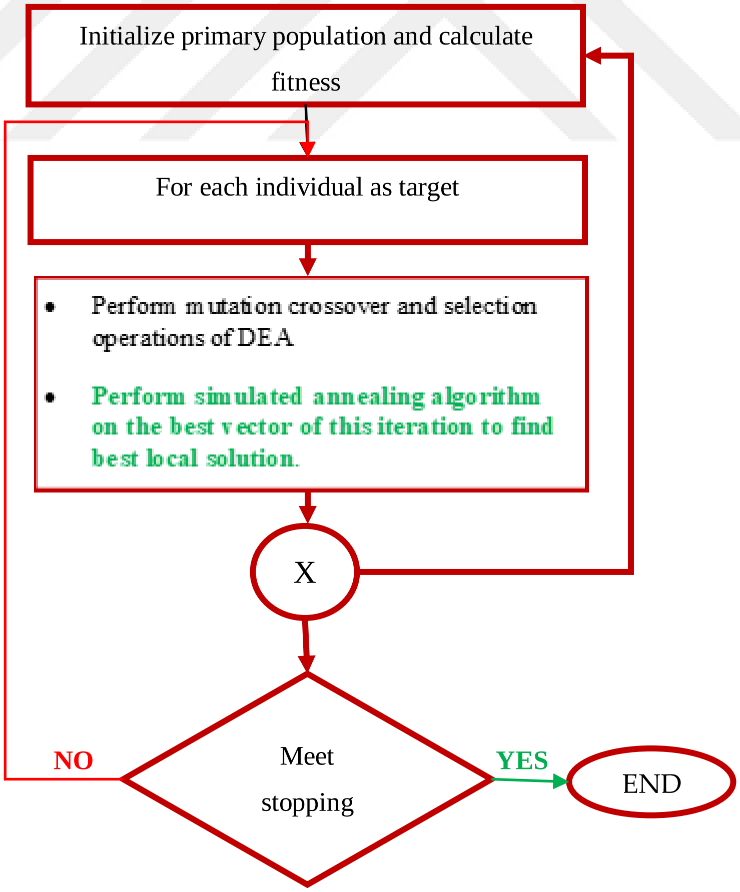


Figure 4.5: Hybridization between DEA and SA

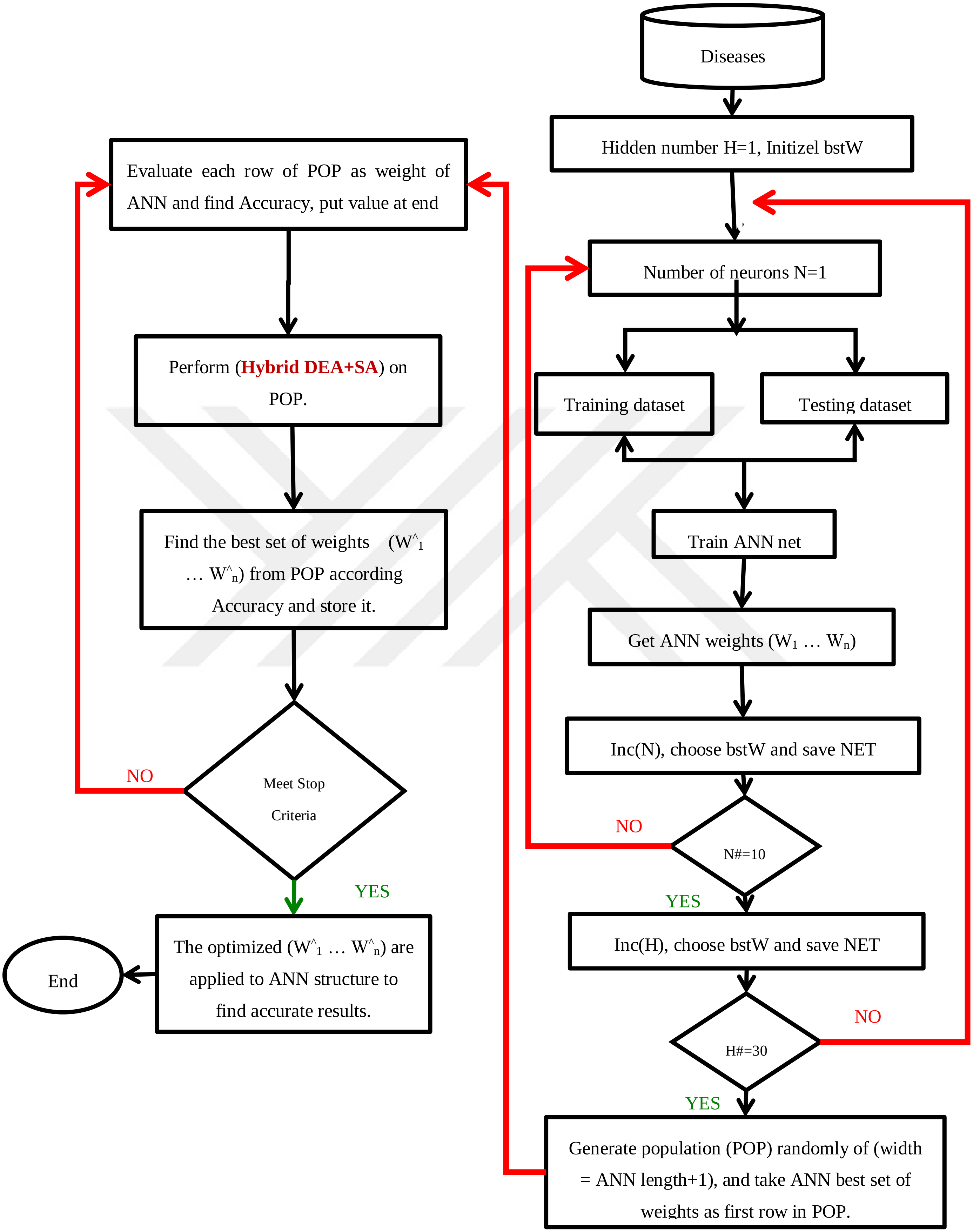


Figure 4.6: Proposal methodology

The parameter settings are illustrated in Table 4.1, Setting for the proposed algorithms TS, SA, GA, DE, and DESA when tested on anemia, PID, and LD.

Table 4.1: Parameter setting for the proposed work

DE		**GA**	
Parameters	Value	Parameters	Value
Number of generations	1000	Number of iterations	1000
size of population	50	size of population	50
Rate of crossover	0.8	Rate of crossover	0.7
		Mutation rate	0.3
TS		**SA**	
Parameters	Value	Parameters	Value
Number of iterations	1000	Number of iterations	1000
		Temp	0.025
		Alpha	0.99
DESA method setting			
DEA part parameter setting		SA part parameters setting	
Parameters	Value	Parameters	Value
Number of generations	1000	Number of iterations	100
size of population	50	Temp	0.025
Rate of crossover	0.8	Alpha	0.99

To effectively evaluate the proposed work ANN_DESA, we tested the three datasets on the ANN hybridized with two trajectory methods SA and TS and with two evolutionary algorithms GA and DE. Finally, we optimized the ANN by hybridization of DE and SA with DESA. In order to measure the diverse classifiers performance relative to our approach, experimental results based on classification accuracy have been obtained. The accuracy was calculated in Eq. (4.6).

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

where TP, TN, FP, and FN are true positive, true negative, false positive, false negative respectively.

4.2.5 Diseases Datasets

The classification and optimization algorithms of ANN are hybridized on the following data sets of disease:

(1) PID

According to the National Institute of Diabetes and Digestive and Kidney Diseases, the data set comprises Pima Indian heritage patients who pregnant females older than 21 years old.

Tuples number: 768

Attributes: 9 (consisting class).

(2) LD

Seven attributes are included in this data set. The results of blood test associated with liver disorders due to alcohol consumption are indicated in the five attributes. The attribute number six is constituted by the drinks number per day.

Patient class and condition (i.e., whether the patient has the disorder or not) are presented in the seventh attribute.

Tuples number: 345

Attributes number: 7 (consisting class attribute)

(3) Anemia

real data set taken from blood laboratory in AL-Anbar health directorate / Iraq [130]. 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. [131].

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 values of attribute are illustrated in Table 4.2.

Table 4.2: Anemia dataset specifications (attributes)

Abbreviations	**Explanation**	**Units**
Age	Patient age	Years
Gender	Patient gender	1: male 2: female
HP	Hemoglobin	(G/dl $\times 10^6$)
RBC	Red blood cell	(Cell/µl $\times 10^3$)
HCT	Hematocrit	(%)
MCV	Mean corpuscular volume	(fL)
MCH	Mean corpuscular hemoglobin	(pg)
MCHC	Mean corpuscular hemoglobin concentration	(G/dl $\times 10^6$)
WBC	White blood cell	(Cell/µl $\times 10^3$)
PLT	Platelets	(Cell/µl $\times 10^3$)

In general, a brief explanation of the data set which is used at the study is shown in Table 4.3.

Table 4.3: Study benchmarks Description

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

4.3 RESULTS

In the context of the study, three disease datasets, namely, PID and LD from the UCI repository and anemia dataset from the AL-Anbar health provenance laboratories, were tested with the two trajectory algorithms SA and TS, two evolutionary algorithms GA and DE, and finally with the proposed method DESA. Table 4.4 illustrates the empirical results.

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**
PID	78.91	80.86	80.86	82.03	80.86	**84**
LDs	75.65	78.84	78.26	77.68	78.26	**79.8**

Obviously, the results of DESA were superior to those of all the other algorithms. Hence, we concluded that the trade-off between the exploration and the exploitation can widely benefit the exploration of search space and intensive exploitation of the greatest capable solutions within local area of the search and balance between the two operations.

Statistical t-test was performed after 25 runs of ANN against ANN+DESA to test the statistical significance of differences. The t-test statistics were transformed into a conditional probability which is named as p-value. The query "If the null hypothesis is true, then what is the probability of observing the current data or the more extreme data?" is answered by the p-value Eq. (3.7) in section 3.3.4. The p-value of the null hypotheses was less than 0.0001, and the difference was extremely significant. The test values are listed in Table 4.5.

Table 4.5: Statistical test of the three medical datasets PID, LD, and anemia between ANN and ANN + DESA

	PIMA	**LD**	**Anemia**
Mean difference	0.03684000	0.002676	0.010216

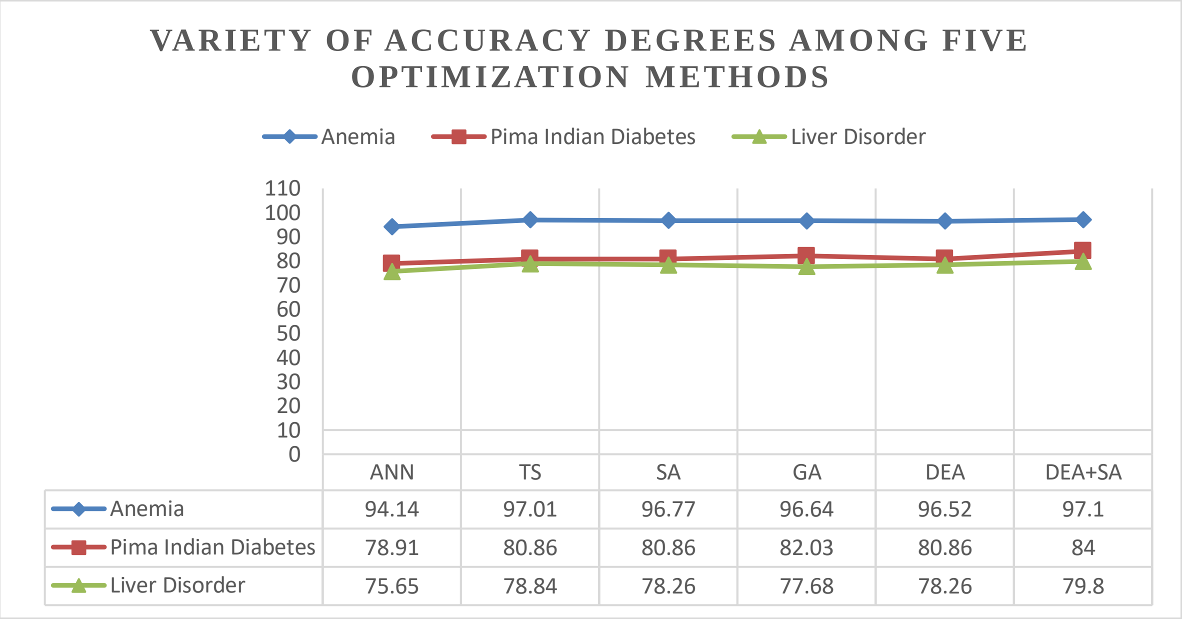
T score	14.1156	36.5694	11.8647
standard error of difference	0.005	0.002	0.001
Difference verdict	Significant	Significant	Significant

4.4 DISCUSSION

As ANN structure, the number of hidden layers and neurons number in every layer are critical in ANN leaning. A suitable ANN structure was built by selecting the finest hidden layers number and the suitable number of neurons in each layer to obtain a structure that gives the best accuracy in medical dataset classification. Boosting the ANN model complexity by raising the hidden layers number and neurons number may lead to overfitting. Metaheuristic algorithms normally solve the ANN overfitting problem. However, in the context of this study, ANN exclusively pushes for a potential high overfitting in the training phase. Therefore, ANN requires a metaheuristic method that can explore the search space to fulfill the diversification and simultaneously perform intensive exploitation for the most promising solutions in the local areas.

The DESA method represents an example of a balance between the diversification and the intensification. The exploration feature of the DE and the exploitation feature of the SA algorithm were combined to reach the global optimum solution.

When hybridizing two metaheuristic algorithms among a list of metaheuristic algorithms, selecting two high-accuracy algorithms is not mandatory. In anemia as an example, we noted the accuracy of TS to be greater than that of SA and the accuracy of GA to be greater than that of DE despite that best hybridization achieved between DE and SA Table 4.4, and **Figure 4.7**. This hybridization combined the two algorithms perfectly and performed more effectively than all the other algorithms.



	ANN	TS	SA	GA	DEA	DEA+SA
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

Figure 4.7: Accuracy degrees of anemia, PID, and LDs tested with optimization methods

All the metaheuristic algorithms chosen in this study recorded an enhancement when hybridized with the ANN despite a low improvement. The difference in enhancement was not large but the accuracy results of the classification were high, considering literature and the study in chapter 3, This result was achieved because we have already relied on the neural network model to find the greatest value for accuracy.

4.5 CONCLUSIONS

This study aimed to propose a metaheuristic method to facilitate balance between exploration and exploitation and consequently enhance medical dataset classification. The hidden layers number with the suitable amount of neurons in each layer can affect ANN learning. Preparing an ANN structure by selecting a complex structure to achieve high accuracy can benefit metaheuristic algorithm efficiency for searching the global optimum. When combining two metaheuristic algorithms to formulate a new improved algorithm for any reason, choosing the algorithm with highly accurate result between the two is not mandatory. Instead, empirical tests can be performed to find convenience. Our approach can be adopted in future studies to solve various problems.

5. CONCLUSION

This chapter presents the overall conclusion of the major advancements corresponding to the work reported in this thesis. The thesis integration employs the mapping of research questions, research objective, and contributions. The overall conclusion of this work is obtained by using standard and real-world data, i.e., the proposed context of this thesis.

In the first study, ANN is hybridized by using three optimization algorithms (PSO, FW, and GA). The algorithms are adopted for ANN hybridization with diverse layer structures and different numbers of metaheuristic iterations. The parameters are also tuned in this scheme. Unlike most of the other approaches in the literature, PSO is competitive on the five utilized medical datasets. The proposed methodology is tested with the big-data classification RNA-Seq that contains thousands of attributes. The experiment on big-data RNA-Seq classification is a practical application of NFLT. This theorem states that no universal optimization algorithm is existed for the whole optimization problems. Our approach can be adopted in future studies when solving for various medical problems.

In the second study, the hybridization between differential evolution (DE) and simulating annealing (SA) is conducted to enhance ANN. The proposed ANN_DESA method, which balances exploration (diversification) and explanation (intensification), is tested on the real medical dataset classification of anemia cases. The proposed methodology begins with the selection of the best ANN structure (i.e., best hidden layers number and best neurons number in each layer). The classification process involves the hybridization of DEA and SA to improve classification accuracy. The statistical t-test of ANN against ANN+DEASA is performed to establish statistically significant differences, and p-values for the null hypotheses are designated. The proposed metaheuristic method facilitated the balance between exploration and exploitation. An enhanced medical dataset classification was the aim of this study. The hidden layers number and the neurons number in each layer can affect ANN learning. In this thesis, ANN is constructed as a complex structure with high accuracy. Consequently, the highly accurate results can further lead to metaheuristic algorithm efficiency when searching for the global optimum. Two metaheuristic algorithms are combined to formulate a new and better algorithm that can handle any problem domain. Nonetheless, choosing between two highly accurate results is not mandatory; instead, empirical tests can be conducted for convenience. For example, in this study’s context, SA obtained lower values compared with TS. Furthermore, the hybridized GA–ANN is superior to DE. However, after combining DE and SA to obtain ANN+DEASA, a remarkable result is observed. Even if no scientific or mathematical rules govern the implementation of metaheuristic hybridization except for those in the empirical tests, the topic is still an interesting field of research, particularly for multi-problem domains such as those in the medical field. Our approach can be adopted in future studies to solve various 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] I. Salman, O. N. Ucan, O. Bayat, and K. Shaker, “Impact of metaheuristic iteration on artificial neural network structure in medical data,” *Processes*, vol. 6, no. 5, 2018.

[82] 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.

[83] 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.

[84] 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.

[85] 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.

[86] 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.

[87] 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.

[88] 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.

[89] 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.

[90] R. Storn and K. PRICE, “Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces,” pp. 341–359, 1997.

[91] 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.

[92] O. S. Soliman and E. Aboelhamd, “Classification of Breast Cancer using Differential

Evolution and LeastSquares ABSTRACT :,” vol. 3, no. 2, pp. 155–161, 2014.

[93] I. De Falco, “Differential Evolution for automatic rule extraction from medical databases,” *Appl. Soft Comput. J.*, vol. 13, no. 2, pp. 1265–1283, 2013.

[94] H. A. Abbass, “An evolutionary artificial neural networks approach for breast cancer diagnosis,” *Artif Intell Med*, vol. 25, no. 3, pp. 265–281, 2002.

[95] J. Kennedy and R. Eberhart, “Particle swarm optimization,” *Neural Networks, 1995. Proceedings., IEEE Int. Conf.*, vol. 4, pp. 1942–1948 vol.4, 1995.

[96] 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.

[97] 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.

[98] 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.

[99] Y. Tan, “Fireworks Algorithm,” pp. 247–262, 2015.

[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] H. H. Örkcü, M. Đ. 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.

[105] S. Mandal and I. Banerjee, “Cancer Classification Using Neural Network,” vol. 3, no. 7, pp.

172–178, 2015.

[106] 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.

[107] 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.

[108] 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].

[109] 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.

[110] M. Maddouri and M. Elloumi, “A data mining approach based on machine learning techniques to classify biological sequences,” vol. 15, pp. 217–219, 2002.

[111] 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.

[112] M. R. G. S. C. P. M. Pardalos, “Classification and Characterization of Gene Expression Data with Generalized Eigenvalues,” pp. 533–545, 2009.

[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] W.-H. A. W.-H. Au and K. C. C. Chan, “Classification with degree of membership: a fuzzy approach,” *Proc. 2001 IEEE Int. Conf. Data Min.*, pp. 24–25, 2001.

[117] L. P., “Classification Based on Fuzzy Robust PCA Algorithms and Similarity Classifier,” *Expert Syst. Appl.*, vol. 36, no. 4, pp. 7463–7468, 2009.

[118] Y. Lee, O. L. Mangasarian, and W. D. Street, “SSVM : A Smooth Support Vector Machine

for Classification," no. 5, pp. 1–22.

[119] J. Yang, Y. Ma, X. Zhang, S. Li, and Y. Zhang, "A Minimum Spanning Tree-Based Method for Initializing the K-Means Clustering Algorithm," vol. 11, no. 1, pp. 13–17, 2017.

[120] I. Rustempasic and M. Can, "Diagnosis of Parkinson ' s Disease using Principal Component Analysis and Boosting Committee Machines," *SouthEast Eur. J. Soft Comput.*, vol. 2, no. 1, pp. 102–109, 2013.

[121] 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.

[122] E. H. Elshami and A. M. . Alhalees, "Automated Diagnosis of Thalassemia Based on DataMining Classifiers," no. June 2012, 2015.

[123] A. Yilmaz, "A FUZZY EXPERT SYSTEM DESIGN FOR," no. October, 2013.

[124] 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.

[125] N. Amin and A. Habib, "Comparison of Different Classification Techniques Using WEKA for Hematological Data," no. 3, pp. 55–61, 2015.

[126] 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.

[127] 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.

[128] V. Kachitvichyanukul, "Comparison of Three Evolutionary Algorithms :," vol. 11, no. 3, pp. 215–223, 2012.

[129] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, "Optimization by Simulated Annealing," vol. 220, no. 4598, pp. 671–680, 1983.

[130] "BAnemia." .

[131] 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.

جمهورية العراق
محافظة الانبار
دائرة صحة الأنبار
قسم الإمور الفنية
شعبة المختبرات
العدد : ١٢٠٢٠٥
التاريخ : ٢٠١/ ١٢/ 2017

((محاربة الفساد خطوة مهمة لبناء المستقبل))

دائرة صحة الانبار
Al-anbar health directorate

الى / السيد احسان سلمان جاسم

م/تسهيل مهمة

استنادا الى طلب (السيد احسان سلمان جاسم / دكتوراه حاسبات واستنادا الى موافقة اللجنة العلمية والاخلاقية في دائرتنا .
بخصوص تسهيل مهمة والتزود ببيانات فحصوات طبية لمرضى فقر الدم (Anemia) من مراكز الفحص لتطوير نماذج رياضية وخوارزميات تحسين فائقة لتصنيف وتشخيص الامراض من خلال بيانات الفحوصات اسم البحث العلمي هو :
(metaheuristic algorithms to enhance medical data classification)

نود اعلامكم بانه لا مانع لدينا من تسهيل مهمة الباحثين السماح لهم باستخدام البيانات لاغراض البحث الاكاديمي على ان لا تتحمل دائرتنا اي تبعات مالية من ذلك.

للتفضل بالاطلاع .... مع الاحترام

د . الدكتور
عبد الستار كاظم لواص
المدير العام
٢٨ / ١٢ / 2017

نسخة منه الى :-
قسم الامور الفنية\ شعبة المختبرات /مع الاوليات

Republic of Iraq
Anbar province
Al-Anbar health directorate
Technical department / laboratories section
No.: 14305
Date: 28/12/2017

To / Mr. Ihsan Salman Jasim

Task Facilitation

According to request of (Mr. Ihsan Salman Jasim/ PhD in computer), and according to our scientific and ethical committee agreement about task facilitation to have medical tests for (Anemia cases) from laboratories to develop metaheuristic algorithms to enhance medical data classification and diagnoses.

In order to accomplish his research:

(Hybrid metaheuristic algorithms to enhance medical data classification)

We confirm that we don't mind facilitating researcher task to support him with (Anemia) data set and allow him to use this data set for academic research purpose without any financial supporting.

For your kind information …………. with respect

Dr. Abdulsattar Kazim Lawas

General Manager

28/12/2017

Copy to:

- Technical department / Laboratories section / wit Precedents ….