REPUBLIC OF TURKEY

YILDIZ TECHNICAL UNIVERSITY

GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

MATHEMATICAL MODELING TO PREDICT THE ANEMIA BASED ON MEDICAL DATA

Arshed A. AHMAD

DOCTOR OF PHILOSOPHY THESIS

Department of Mathematics

Program of Mathematics

Advisor

Prof. Dr. Murat SARI

January, 2020

REPUBLIC OF TURKEY

YILDIZ TECHNICAL UNIVERSITY

GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

MATHEMATICAL MODELING TO PREDICT THE ANEMIA BASED ON MEDICAL DATA

A thesis submitted by **Arshed A. AHMAD** in partial fulfillment of the requirements for the degree of **DOCTOR OF PHILOSOPHY** is approved by the committee on 31.01.2020 in Department of Mathematics, Program of Mathematics.

Prof. Dr. Murat SARI
Yildiz Technical University
Advisor

Approved By the Examining Committee

Prof. Dr. Murat SARI, Advisor
Yildiz Technical University _____

Prof. Dr. İdris DAĞ, Member
Eskisehir Osmangazi University _____

Prof. Dr. Bayram Ali ERSOY, Member
Yildiz Technical University _____

Prof. Dr. Dursun IRK, Member
Eskisehir Osmangazi University _____

Assoc. Prof. Dr. Özgür YILDIRIM, Member
Yildiz Technical University _____

I hereby declare that I have obtained the required legal permissions during data collection and exploitation procedures, that I have made the in-text citations and cited the references properly, that I have not falsified and/or fabricated research data and results of the study and that I have abided by the principles of the scientific research and ethics during my Thesis Study under the title of "MATHEMATICAL MODELING TO PREDICT THE ANEMIA BASED ON MEDICAL DATA" supervised by, Prof. Dr. Murat SARI. In the case of a discovery of false statement, I am to acknowledge any legal consequence.

Arshed A. AHMAD

Signature

I would like to dedicate my thesis to my father and my mother, as well as to my family, who endured this long study period with me, always offering patience, support, and love.

ACKNOWLEDGEMENTS

Praise be to Allah for giving me strength and guidance 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. Murat SARI for his valuable guidance and standing beside me throughout my studies and his support, generosity and freedom throughout the entire research and thesis writing. 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 Yildiz Technical University, Istanbul, Turkey.

Also, my sincere thanks go to my family: my parents for their immense patience and unconditional support and encouragement throughout my life, and to my wife, Raya, my children, Mira and Aram, my brothers and sisters for their support and encouragement throughout my thesis. I would like to thank my friends who give me constructive comments and warm encouragement.

Last but not the least, my sincere thanks also go to the Iraqi Ministry of Higher Education and Scientific Research, and Diyala University for giving me this scholarship which enabled me to get my Ph.D.

Arshed A. AHMAD

TABLE OF CONTENTS

LIST OF SYMBOLS

y_i	Dependent Observations
x_i	Independent Observations
B	Regression Coefficients
ϵ	Unobserved Random Variable
k	Predictor Variables
$\hat{\mathbf{y}}$	Estimated Value
n	The Number of Observed Data
$\bar{y}$	The Mean of Dependent Observations
R^2	Determination of the Coefficient
e_i	Vector of Residuals
t	Iteration Number
ω	Weight Parameter
c_1, c_2	Acceleration Coefficients
r_1, r_2	Random Numbers
$V_i{}^t$	Position of individual i at iteration t
$X_i{}^t$	Velocity of individual i at iteration t
P_{best}	Best Local Value of Each Particle
G_{best}	Best Value of Swarm
$Pbest$	Best Value

$lbest$	Best Location
$Gbest$	Global Best
$P-Value$	Probability Value
$F-Stat$	*F*–test statistics
$t-Stat$	*t*–test statistics

LIST OF ABBREVIATIONS

HB	Hemoglobin
RBC	Red Blood Cell
MCV	Mean Corpuscular Volume
HCT	Hematocrit
MCHC	Mean Corpuscular Hemoglobin Concentration
RDW	Red Cell Distribution Width
WBC	White Blood Cell
MCH	Mean Corpuscular Hemoglobin
PLT	Platelets
PSO	Particle Swarm Optimization
WHO	World Health Organization
MRA	Multiple Regression Analysis
MLR	Multiple Linear Regression
SST	Sum of Squares Total
SSR	Sum of Squares Regression
SSE	Sum of Squares Error
MSE	Mean Square Error
RMSE	Root Mean Square Error
SD	Standardized Divation
LSTM	Long Short–Term Memory

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

MATHEMATICAL MODELING TO PREDICT THE ANEMIA BASED ON MEDICAL DATA

Arshed A. AHMAD

Department of Mathematics

Doctor of Philosophy Thesis

Advisor: Prof. Dr. Murat SARI

Different diseases and diagnostic methods using various tests produced large amounts of complex medical data. Therefore, huge number of patient records in clinical centers, hospitals, and other health institutions have created the need for developed and accurate medical applications to help doctors. Since anemia is one of the most common health problems in recent era, the aim of this thesis is to predict anemia from a population through biomedical variables of individuals (the blood variables, age, and sex) and the anemia types using the currently produced mathematical models. This work is carried out using the dataset consisting of 539 subjects provided from blood laboratories. This thesis basically focuses on mathematical modeling to predict the anemia problem based on medical data. The main problems associated with medical diagnose involve the identification of highly accurate prediction models. For the first step, a mathematical method based on multiple linear regression (MLR) analysis has been applied to a reliable model that investigate if there exists a relation between the anemia and the biomedical variables and to provide the more realistic one. For the second step, a multiple nonlinear regression analysis has been used for a reliable model that research if there exists a mathematical relation between the observational variables and the anemia types. The parameter values produced are all seen to be the optimum values obtained from the multiple regression approaches, to provide the more realistic one. At the last step, optimum medical models based on biomedical variables are produced and an effective technique is used in investigating the optimum parameters of the models. To achieve this, the particle swarm optimization (PSO) algorithm has effectively been applied in predicting the parameters of the mod-

els through the biomedical variables. Optimum values of the parameters produced from the PSO algorithm are used here to obtain more realistic models. The current models have been compared with the other ones and the results have been seen to be better. The models based on the variables and outcomes are expected to serve as a good indicator of disease diagnosis for health providers and planning treatment schedules for their patients. Thus, the study has been seen to be beneficial especially for those are interested in biomedical models arising in various fields of medical science, especially anemia.

Keywords: Anemia, Medical modelling, Mathematical modelling, Regression model, Particle swarm optimization, Nonlinear model.

ÖZET

TIBBİ VERİLERE DAYALI ANEMİYİ TAHMİN ETMEK İÇİN MATEMATİKSEL MODELLEME

Arshed A. AHMAD

Matematik Bölümü

Doktora Tezi

Danışman: Prof. Dr. Murat SARI

Birden fazla test kullanan farklı hastalıklar ve tanı yöntemleri büyük miktarlarda karmaşık tıbbi veriler üretmiştir. Bu nedenle, klinik merkezler, hastaneler ve diğer sağlık kurumlarındaki çok sayıda hasta kaydı, hastanın kritik durumda olup olmadığına bakılmaksızın veya uzaktan takip gerektirmeksizin doktorların ve terapistlerin vakaları araştırmasına yardımcı olmak için gelişmiş ve doğru tıbbi uygulamalara ihtiyaç duymuştur. Anemi günümüzde en sık rastlanan sağlık sorunlarından biri olduğundan, bu tezin amacı bireylerin biyomedikal değişkenlerini (kan değişkenleri, yaş ve cinsiyet) kullanarak anemi olup olmadıklarını bulmak ve halihazırda üretilen matematiksel modelleri kullanarak anemi türünü tahmin etmektir. Bu çalışma, kan laboratuvarlarından sağlanan 539 denekten oluşan veri ile gerçekleştirilmiştir. Bu tez, temel olarak tıbbi verilere dayalı anemi problemini tahmin etmek için matematiksel modellemeye odaklanmaktadır. Tıbbi teşhislerle ilişkili temel problemler, doğru tahmin modellerinin tanımlanmasını içerir. İlk adım için, anemi ve biyomedikal değişkenler arasında bir ilişki olup olmadığını araştıran ve daha gerçekçi olanı sağlayan güvenilir bir model için çoklu doğrusal regresyon analizine dayanan bir matematiksel yöntem uygulanmıştır. İkinci adım için, gözlemsel değişkenler ve anemi türleri arasında bir ilişki olup olmadığını araştıran güvenilir bir model için doğrusal olmayan çoklu regresyon analizi yöntemi kullanılmıştır. Üretilen parametre değerlerinin hepsinin, daha gerçekçi olanı sağlamak için çoklu regresyon yaklaşımlarından elde edilen optimum değerler olduğu görülmektedir. Son adımda, biyomedikal değişkenlere dayanan optimum doğrusal tıbbi model üretilir ve modelin optimum parametrelerinin araştırılmasında etkili bir teknik kullanılır. Bunu başarmak için, parçacık sürüsü optimizasy-

onu (PSO) algoritması, modelin parametrelerini biyomedikal değişkenler aracılığıyla tahmin etmede etkili bir şekilde uygulanmıştır. PSO algoritmasından üretilen parametrelerin optimum değerleri burada daha gerçekçi bir model elde etmek için kullanılır. Mevcut modeller diğer yöntemlerle karşılaştırıldığında; mevcut sonuçların daha iyi olduğu görülmektedir. Değişkenlere ve sonuçlara dayanan modelin, sağlık hizmeti sunanlar açısından hastalık teşhisi için iyi bir araç ve hastalar için doğru tedavi planlaması beklenmektedir. Bu nedenle, çalışmanın özellikle tıp biliminin bir çok farklı alanında ve anemi teşhisinde ortaya çıkan biyomedikal modellerle ilgilenenler için yararlı olacağı görülmüştür.

Anahtar Kelimeler: Anemi, Tıbbi modelleme, Matematiksel modelleme, Regresyon modeli, Parçacık sürü optimizasyonu, Doğrusal olmayan model.

1

INTRODUCTION

This chapter presents an overview of the anemia problems and their solution techniques that have been offered in the literature. Also, the goal of this research and the hypothesis of the thesis are given respectively.

1.1 The Literature Review

A data model is an abstract model that organizes elements of the data and standardizes how they relate to each other and to the properties of real-world entities. A data model is based on data, data relationship, and data constraint. A data model provides the details of information to be stored and is of basic use when the result is the generation of algorithm for an application or the preparation of a functional specification to aid a computer software make decision. Therefore, the study of observational data to discover their relations and to summarize the behavior of data are both understandable and useful for human [1, 2].

In recent years, data has occupied great interest in information systems. This is because existing computers are able to create and store almost unlimited datasets. In fact, database and information technology have grown systematically from primal file processing systems to complicated and powerful database systems [3].

So, a data model focuses on representing the data as a user sees it in the "real world". It serves as a bridge between the concepts that make up the real-world processes and the physical representation of those concepts in a database [3].

Health data is any information related to the physical or mental health conditions of individuals, reproductive outcomes, causes of death, and quality of life of the individual or population. Health information includes clinical metrics along with environmental,

social, economic and behavioral information related to health and wellness. A lot of data is collected, stored, processed and used when individuals interact with healthcare systems. A large collection of such data collected by health providers may be included to that. Increased collection data of patients is a major component of digital health. Thus, medical data refer to health-related information associated with patient care regularly or as a part of a clinical trial program [4, 5].

Medical diagnosis is considered as a very important issue that requires adequate and proper implementation, hence, designing accurate models in this area would be very beneficial in diagnosing process and health providers take advantage of models to diagnose very complicate cases with a large number of patients in time with predefined diagnosing models based the models. An automatic medical diagnosing model is possible to be extremely advantageous by bringing the whole materials, tools, and objectives together in order to make a relation with a diagnosis target.

Medical data analysis and diagnosing diseases as acknowledged discovery is important but hard missions and naturally is based on years of practice of a professional as seen in Saez et al. [6]. As pointed out by Liu et al. [7], initial medical diagnosis model from patients' medical reports in early time has an important meaning for accurate health treatment.

mathematical model is a relationship that includes all variables of a problem. Therefore, it is a description of a system that uses mathematical concepts and language. A model can help to explain a system, to study the effects of different components and to make predictions about the behavior of a system. This process of repeated iteration is a typical modelling project and is one of the most useful aspects of modelling in terms of improving our understanding of how the system works. Note that, a mathematical model depends on the data model. Thus, a key determinant of the potentiality of a given model to help in such measures is the availability of data to parameterize the model. It is therefore important to understand the types of data that are necessary for a modeling project to be successful.

In mathematical modeling, the values of dependent variables depend on the values of independent variables. The dependent variables represent the output whose variation is being studied. The independent variables represent inputs or causes, potential reasons for variation.

It is helpful to divide up the process of a model into four categories of activity: building, studying, testing and use.

In general, defects found at the studying and testing stages are corrected by returning to the building stage. Note that if any changes are made to the model, then the studying and testing stages must be repeated [8].

A pictorial representation of potential routes through the stages of modelling can be seen in Figure 1.1.

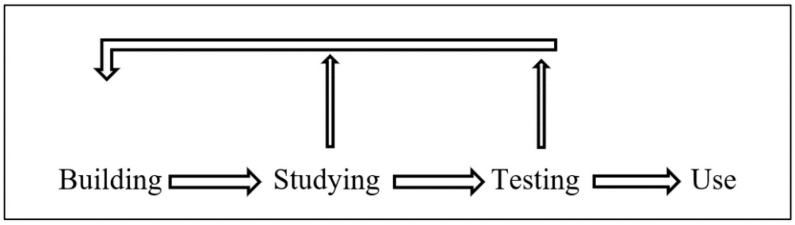


Figure 1.1 Process of modelling

The data model focuses on what data is required and how it should be organized rather than what operations will be performed on the data. A data model is independent of hardware or software constraints. The data model focuses on representing the data as the user sees it in the real world, therefore, it serves as a bridge between the concepts that make up real-world events [8, 9].

The construction progress of mathematical models considered to produce medical outputs is a growing field in medical science. Mathematical models used for various reasons are important equipments to deal with the behavior of a medical problem. Mathematical modeling has been realized to be a fundamentally important tool for the analysis of pathological characteristics. Therefore, finding a response to a medical model with high performance is of major interest and thus the medical model can describe the relationship between the biomedical variables and the diseases. Therefore, the researchers used the mathematical model to predict their problems, especially their medical problems, such as Sari et al. [10] proposed two systems, artificial neural network, and adaptive neuro-fuzzy inference system, to predict the low back pain level. A particle swarm optimization (PSO) and K-means clustering algorithm have been combined to predict tibial rotation pathologies through divided datasets into three clusters by Sari et al. [11]. Also, Sari and Cetiner [12] used the artificial neural networks to predict the effect of physical factors on tibial motion through gender, age,

body mass, and height. Li et al. [13] studied the prediction of the risks of congenital heart disease in pregnant women by using the artificial neural network through hospital-based case-control study. Combining the genetic algorithm with the neural network are predicting the risk of cardiovascular disease by Amma [14]. Sisodia and Sisodia [15] used three machine learning algorithms, Decision Tree, SVM, and Naive Bayes to predict diabetes. Also, Zou et al. [16] adapted the decision tree approach, random forest, and neural network to predict diabetes mellitus.

Also, various researchers used mathematical models to predict anemia and blood disease problems, such as Martínez-Martínez et al. [17] used machine learning techniques to predict the hemoglobin (HB) level in hemodialysis patients. Reymann et al. [18] proposed an algorithm to predict blood glucose levels through support vector regression model. Altrock et al. [19] developed a mathematical model that describes the aging and survival of sickle susceptible and normal RBCs. Implementations of some classification algorithms have been seen through various methods in the literature such as Naïve Bayes, Neural network, Decision Tree (J48), and Support Vector Machine to predict anemia types (chronic anemia, iron deficiency anemia, anemia of renal disease, thalassemia, and aplastic anemia) through Mean Corpuscular Volume (MCV), Hematocrit (HCT), HB, Mean Corpuscular Hemoglobin Concentration (MCHC), Red Cell Distribution Width (RDW) by Abdullah and Al-Asmari [20]. An application of a multilayer perceptron neural network to estimate missing values and predict the degree of post-operative anemia by Yu et al. [21]. Two anemia types, iron deficiency, and thalassemia were investigated with white blood cell (WBC), RBC, HB, HCT, MCV, mean corpuscular hemoglobin (MCH), MCHC, RDW, and PLT by five classification algorithms and a vote algorithm were used by Hasani and Hanani [22]. El-Halees and Shurrab [23] predicted a blood tumor by using three different methods of data mining which are association rules, rule induction, and deep learning and through WBC, RBC, HB, HCT, MCV, MCH, MCHC, RDW, PLT. Two algorithms were examined Hamdi et al. [24], support vector regression and differential evolution algorithms for prediction of continuous blood glucose, their algorithm achieved high prediction accuracy. Tetschke et al. [25] built a mathematical model which has ability to capture the most important features to predict of RBC Count after blood loss through HB, HCT, MCH, RBC. A simple coronary disease prediction model was developed using a gradient boosting

decision tree algorithm by using WBC, RBC, MCHC, HCT, MCV, PLT, HB Meng et al. [26]. Jaiswal et al. [27] suggested machine learning algorithms, Naive Bayes, random forest, and decision tree algorithm for the prediction of anemia disease with HB, RBC, HCT, MCH, MCV.

Multiple regression analysis (MRA) is a statistical tool that predicts the value of a dependent variable based on the multi-independent variables. Thus, once the multiple variables related to a dependent variable are determined, any information about all predictor variables can be realized and used to make more accurate predictions. Therefore, researchers applied regression techniques to predict anemia and blood diseases by models based on blood variables, such as relating soil lead levels to predict children's blood lead levels through a multivariate linear regression model by Lewin et al. [28]. Prediction of anemia was done by Makh et al. [29] in intrauterine growth by applying linear regression analysis. Also, Foster et al. [30] applied the MRA to predict anemia on unenhanced computed tomography of the thorax through HB, HCT. Vincent et al. [31] built a multivariable logistic regression model to find out chemotherapy-induced anemia in patients with non-advanced cell lung cancer through HB testing. Also, Schneider et al. [32] applied a multiple regression for identifying risk factors related to anemia and iron deficiency in a sample of children. Lee et al. [33] used a simple regression analysis to identify the relationship between HB or HCT level and dural sinus density. The development of a set of 14 models with a genetic risk score, a set of these models were used by Milton et al. [34] to forecast fetal hemoglobin in patients with sickle cell anemia, the association was tested using a linear regression model. Determining risk factors for anemia in children depending on the hemoglobin concentration in the blood were determined Dey and Raheem [35] using a multilevel regression model. Building a linear regression model was built by Hsieh et al. [36] through pulse transit time to predict blood pressure. Chen and Miaou [37] proposed an anemia testing approach by applying a Kalman filter and a regression method. Determinants of childhood anemia was evaluated by Habyarimana et al. [38] by applying the quantile regression model and the test of the HB. Aishah et al. [39] verified the relation of fasting blood glucose, cholesterol, and blood pressure levels in healthy subjects and applied the MLR approach.

The PSO is a randomized, population-based method that helps with optimization prob-

lems. The method works with a set of possible solutions and constraints on an optimization problem. The optimization problem must have a target status then the algorithm runs to solve the problem and provide the best values. Also, it is used on the mathematical models to find the best parameters for the model. Therefore, researchers apply the PSO to improve the efficiency of the models that work to predict anemia and blood diseases. Moreover, the PSO was used to improve the simultaneous selection of the parameters for the calibration of the model using the support vector regression method that estimates blood glucose concentrations [40]. Back-propagation was considered the back-propagation neural network at first, then the PSO based back-propagation networks were applied by Sharma et al. [41] to diagnose the anemia in pregnant women. Blood glucose detection was done by Dai et al. [42] through two artificial neural networks were used as a basic structure of the PSO-ANN model.

Along the literature survey of disease knowledge discoveries, many mathematical models have been tested through various methods and promising results have been obtained. However, investigations in disease prediction are still an open field because of several reasons like:

- There are always new diseases and new tests to discover those diseases.

- Most of the prediction of diseases have not reached the saturation point.

- Scientists always develop new mathematical models and optimization algorithms that give more accurate solutions.

In this study, a mathematical method based on multiple regression analysis has been applied to reliable models that investigate whether there exists a relation between the anemia types and the biomedical variables or not.

1.2 Objectives of the Thesis

This thesis aims at investigating the performance of the MRA and the PSO in order to obtain optimal parameters of the model and at having a capable model representing anemia problems through blood variables, sex, and age.

To achieve this major aim, two objectives are outlined:

1. To derive a new mathematical model to study the effect of the blood variables, sex, and age on the anemia types through a large group of the blood variables.

2. To accurately estimate the parameters of the model.

1.3 Hypothesis of the Thesis

The proposed medical models can be properly applied to identify anemia types through the observational variables. Medical models are able to produce very accurate results to be a good guide for the diagnosis of the anemia types to health providers and planning treatment schedules for their patients.

1.4 Overview of the Thesis

This thesis consists of six chapters. Chapter 1 presents literature review, objectives, and hypothesis of the thesis. The remaining contents are organized as follows:
Chapter 2 describes the problem and summarizes the conventional methods which were applied to the problem. The applied techniques to models were reviewed.
Chapter 3 proposes a multiple linear regression model which is produced through biomedical information to predict the anemia. This prediction has been made by applying the MRA to a mathematical model. The study is conducted in terms of data consisting of 539 subjects provided from blood laboratories. The produced results based on the model were compared. Finally, the linear regression model has been analyzed and discussed.
Chapter 4 presents the details of multiple nonlinear regression analysis used in the model that investigate whether there exists a relation between the anemia and the biomedical variables or not. This work has been carried out in terms of the data in a similar way of Chapter 3. The model results of two rival methods were compared. Finally, the model has been analyzed and discussed.
Chapter 5 focuses on predicting the anemia through biomedical variables by using the optimum models. To achieve this, the particle swarm optimization algorithm has effectively been applied in predicting the parameters of the model through the biomedical variables. The study was conducted in terms of the data in a similar way of Chapter 3. Finally, the models have been analyzed through the optimum values of the parameters produced from the PSO algorithm and discussed.
Some final remarks and recommendations were reported in Chapter 6.

2

BASIC CONCEPTS

2.1 Introduction

This chapter displays a general review of medical 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. This chapter especially focuses on mathematical models to identify an appropriate technique for solving the problem.
The following sections of this chapter are arranged as follows: Section 2.2 provides a brief description of anemia problems and blood variables. An extensive review of techniques applied to the anemia prediction is presented in Section 2.3.

2.2 Anemia Problems

Anemia is defined clinically as hemoglobin value that is below the appropriate reference range for an individual. This decrease in the hemoglobin level leads to decreased oxygen delivery to organs of a body and therefore appears in the symptoms of a headache, fatigue, inability to focus, attention, weakness, exhaustion, chest pain, cold hands, and feet. As signified in the literature [43, 44, 45, 46], the anemia was initially thought to be associated primarily with the infectious, inflammatory diseases. Also, it is a lower hemoglobin level below the normal limits determined by the World Health Organization (WHO) [43]. As pointed out by Hébert et al. [44], anemia is one of the most common cases among blood diseases worldwide. There are many types of anemia. Depending on the types, the symptoms of anemia can range from short episodes to chronic conditions. Each type of anemia produces a different case, ranging from moderate to severe and each has its own causes. Anemia can be either

temporarily or long-term disease.

2.2.1 The Literature Review

In the current study, the data for each subject readings of blood variables are Hemoglobin (HB), Red Blood Cells (RBC), Mean Corpuscular Hemoglobin (MCH), White Blood Cell (WBC), Hematocrit (HCT), Mean Corpuscular Hemoglobin Concentration (MCHC), Platelets (PLT), Mean Corpuscular Volume (MCV) and sex and age. In addition, the anemia types in this study are iron deficiency anemia (1), deficiency vitamin B12 (2), thalassemia (3), sickle cell (4) and spherocytosis (5).
The corresponding blood variables can be briefly introduced as follows. The HB is a portable protein inside the RBC and contains iron atoms, and that carries oxygen from the lungs to the body's tissues and returns carbon dioxide from the tissues back to the lungs. The RBCs are concave cells are useless nucleus contains the HB. The MCH is the calculated value derived from the HB measurement and a number of red cells. The WBCs are the cells of the immune system that are involved in protecting the body against infectious disease. The HCT is percentage of the RBCs volume of total blood volume. The MCHC is the calculated concentration of HB in a specific volume of RBC. The PLT is an irregular, disc-shaped element in the blood that assists in blood clotting. The PLTs are usually classed as blood cells as well. Average size of the red cells in a sample is measured by the MCV. The other biophysical variables, sex and age, are considered. Because natural HB in the body varies from male to female, and thus male: 1, female: 2. Yet, natural HB in the body varies according to age [43, 44].
In the literature, many studies were carried out [47, 48, 49, 50, 51] by using relatively less number of input variables to predict the type of anemia. The methods used in the corresponding studies produced relatively less accurate results. For the blood variables, HB, RBC, MCV, MCH, and Red Cell Distribution Width (RCDW); a study were carried out by Sirachainan et al. [47] to create a mathematical model identifying iron deficiency anemia. They found out a model for detecting beta thalassemia carriers by using the MCV and MCH [48]. Jimnez [49] used the RBC, HB, and HCT for diagnostic value of the common blood disease tests in the distinction between thalassemia and anemia due to iron deficiency. Another researcher [50] considered the MCV, MCH, HCT, and HB exploring the relationship between iron deficiency anemia and academic

achievement third-grade high school female students. Piplani et al [51] used the HB, RBC, MCV, and MCH to assess the validity of 12 different indices to distinguish beta thalassemia trait from the iron deficiency. Despite all those pioneering advances in these fields, the corresponding studies used a relatively limited number of blood variables or a very few numbers of anemia types, usually considered beta thalassemia or iron deficiency anemia.

2.3 The Methods

A medical problem of frequently encountered is that of having a set of data produced from medical analysis, so they are normally too large to derive a mathematical model and to define a set of parameters that characterize the model. In this section, the following techniques are given for solving the medical problem.

2.3.1 Multiple Regression Analysis

The MRA is a useful statistical process that can be used to determine the level of influence of some independent variables on dependent variables and to estimate relationships between the variables. Also, it is a powerful technique used to predict the unknown value from two or more known variables. More specifically, the MRA helps a person understands how to change the typical value of a dependent variable when changing one of the independent variables, while the other independent variables are installed. Multiple regression model allows to analyze the relative effects of these independent or expected variables on the dependent variable and these complex datasets often lead to false conclusions if they are not correctly analyzed [52, 53, 54]. Regression analysis entered social science through the work of Legendre in 1805 and Karl Gauss in 1809. The first form of regression was the least-squares method. Gauss issued another development of the least-squares theory in 1821 with a version of the Gauss-Markov theorem [55, 56]. Galton invented the term regression in the 19th century to depict a biological phenomenon [55, 56].

For Galton, the term regression had only biological meaning, but later, Yule and Pearson edited Galton's work to a more general statistical background, so, Pearson used multiple regression for the first time, 1908, to learn more about the relationship among

several independent variables and a dependent variable [57, 58].

Since multiple regression is a neural network without a hidden layer only, the input and output layer, a regression model can be considered as consisting of just a single neuron. So, regression model, each input is related to each output, in this case there is only a single output, as seen in Figure 2.1 [59].

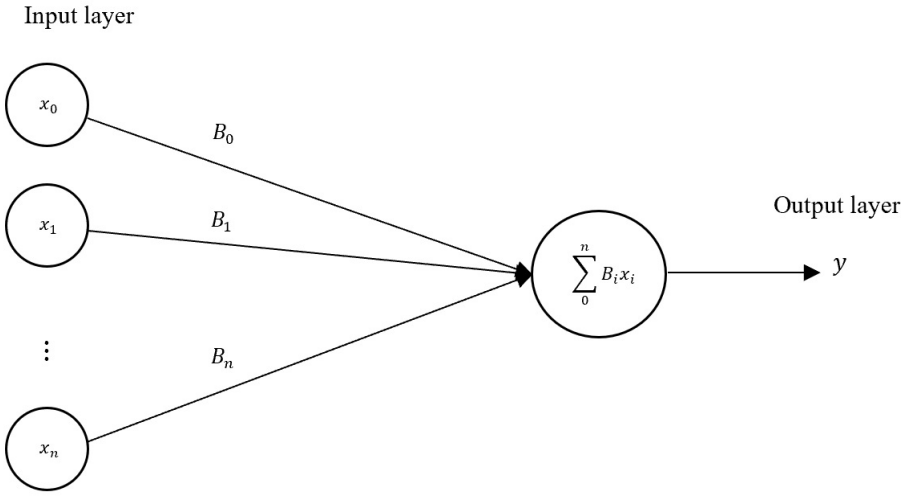


Figure 2.1 Multiple regression is a single-layer neural network

Many researchers have considered the MRA to deal with different problems such as a simple model for weather predicting through the parameters of weather [60], development of a model a dynamic manufacturing system for reducing the gap between theory and real-time data of the system [61], evaluating the energy performance of commercial buildings and to predict any possibility for energy consumption decrease through developing energy consumption indicators for the buildings [62], actual performance of the proton exchange membrane through three temperatures, four flow rates, and two flow patterns [63], hydrogen storage on MgeH2 and LiNH2 under different temperature [64], and wind turbine power curve [65].

Most regression models propose that Y_i is a function of X_i and B, with ϵ representing random statistical noise:

$$Y_i = f(X_i, B) + \epsilon. \tag{2.1}$$

Estimating the function $f(X_i, B)$ that fits with the data is the goal of the researcher. Therefore, we must specify the shape of the function f. Sometimes, the form of this function depends on knowing the relationship between Y_i and X_i, so, a suitable form is chosen for f.

Once the researcher determines the model, various forms of regression analysis provide tools to estimate the parameters B, such as the least-squares to find the value of B by minimizing the sum of the square errors, can be represented as follows:

$$\sum_i (Y_i - f(X_i, B))^2. \tag{2.2}$$

2.3.1.1 Multiple Linear Regression Analysis

In linear regression, relationships are represented by using linear prediction functions that estimate unknown model parameters from the data. Linear regression focuses on the probability distribution of the response in the light of prediction values, rather than the common probability distribution of all these variables. It was the first type of regression analysis to be fully studied and widely used in practical applications. Also, the models that are linearly dependent on their unknown parameters are easier than nonlinear models associated with their parameters and because the statistical properties for resulting estimators are easier to identify.

A linear regression model that contains more than one predictor variable is called a multiple linear regression model. A MLR model with k predictor variables and independent observations

$$\mathbf{y} = B_0 + B_1 x_1 + B_2 x_2 + ... + B_k x_k + \epsilon = B_0 + \sum_{i=1}^{k} B_i x_i + \epsilon. \tag{2.3}$$

The observations recorded for each of these n levels can be expressed in the following way

$$\begin{aligned}
y_1 &= B_0 + B_1 x_{11} + B_2 x_{12} + ... + B_k x_{1k} + \epsilon_1 \\
y_2 &= B_0 + B_1 x_{21} + B_2 x_{22} + ... + B_k x_{2k} + \epsilon_2 \\
&\vdots \\
y_i &= B_0 + B_1 x_{i1} + B_2 x_{i2} + ... + B_k x_{ik} + \epsilon_i \\
&\vdots \\
y_n &= B_0 + B_1 x_{n1} + B_2 x_{n2} + ... + B_k x_{nk} + \epsilon_n.
\end{aligned} \tag{2.4}$$

The dependent observations $y_1, y_2, ..., y_n$, and the independent observations x_1, x_2 $, ..., x_k$, have n levels. Then x_{ij} represents the ith level of the jth predictor variable, x_j.

System (2.4) can be represented as follows:

$$\mathbf{y} = \mathbf{B}\mathbf{X} + \epsilon, \tag{2.5}$$

Here

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix}, \mathbf{B} = \begin{bmatrix} B_0 \\ B_1 \\ \vdots \\ B_k \end{bmatrix}, \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \tag{2.6}$$

where **y**,**X**,**B** and ϵ stand for the observations, the regression coefficients and an unobserved random variable that adds noise to the linear relationship between the dependent variable and regressors, respectively. In matrix notation, these equations can be written as:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix} \begin{bmatrix} B_0 \\ B_1 \\ \vdots \\ B_k \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}. \tag{2.7}$$

To obtain the regression model, **B** should be known. Therefore, **B** is estimated by using the least square estimates as follows

$$\hat{\mathbf{B}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}, \tag{2.8}$$

where $\mathbf{X}^T$ represents the transpose of the matrix **X** while $(\mathbf{X}^T\mathbf{X})^{-1}$ represents inverse of the matrix $(\mathbf{X}^T\mathbf{X})$. Knowing the estimate $\hat{\mathbf{B}}$, the MLR model can now be expressed as [66, 67]

$$\hat{\mathbf{y}} = \hat{\mathbf{B}}\mathbf{X}, \tag{2.9}$$

where $\hat{\mathbf{y}}$ is the estimated value for **y** from the regression.

2.3.1.2 Multiple Nonlinear Regression Analysis

The multiple regression approach has ability to determine the relative effectiveness of one or more variables of the model. In multiple regression, the data is used to describe a relationship between the state variables for the model of interest. Nonlinear regression analysis model can then be given as

$$\mathbf{y} = f(\mathbf{X},\mathbf{B}) + \epsilon, \tag{2.10}$$

where $\mathbf{y}, \mathbf{X}, \mathbf{B}, f()$ and ϵ indicate the observations, the vector of the regression coefficients, the known nonlinear regression function and the unobserved random variable that adds noise to the nonlinear relationship between the dependent variable and the regressors, respectively.

Nonlinear least squares are in the form of least squares analysis used to fit a set of observations with a model that is nonlinear in unknown parameters. The basis of the method is to approximate the model by a linear one and to refine the parameters by successive iterations [68, 69]. First, let

$$y_i = f(X_i, \mathbf{B}) + \epsilon_i, 1 \leq i \leq n, \tag{2.11}$$

and

$$Q = \sum_{i=1}^{k} (y_i - f(X_i, \mathbf{B}))^2. \tag{2.12}$$

In order to find

$$\hat{B} = \arg\min_{B} Q, \tag{2.13}$$

first each of the partial derivatives of Q is found with respect to B_j. Then, each of the partial derivatives is taken to be equal to 0 and the parameters B_k are replaced by $\hat{B}_k, 0 \leq k \leq n$. The functions to be found are nonlinear in the estimates $\hat{B}_k$.

The regression analysis uses the optimization to estimate the parameters of the model by minimizing the sum of the square error. So, the optimization involves minimizing some form of summed squared deviations between the data and the fitted model. This

assumes that a mathematical model has been selected. So, thus, a form of optimization could be considered as the best way of selecting a suitable model [70, 71].

2.3.1.3 Sums of Squares

Sum of squares is a statistical tool used to determine the dispersion of data as well as the suitability of data in the regression analysis model. The sum of squares got its name because it is calculated by finding the sum of squared differences. Therefore, the sum of the smaller squares indicates a fitting model where there is less variation in the data.

The three main types of sum of squares are the sum of squares total (SST), the sum of squares regression (SSR), and the sum of square errors (SSE; also known as the residual sum of squares).

The SST is a variation in the values of a dependent variable from the sample mean of the dependent variable. Basically, the SST determines the overall variance in a sample (see Figure 2.2) and calculated by

$$SST = \sum_{j=1}^{n} (y_j - \bar{y})^2. \tag{2.14}$$

where y_j and $\bar{y}$ indicate the dependent observations and the mean of dependent observations, respectively.

The SSR describes how extent the regression model represents the fit data; therefore, it indicates how good the regression model in explaining the data. The formula for computing the SSR is (see Figure 2.2):

$$SSR = \sum_{j=1}^{n} (\hat{y}_j - \bar{y})^2. \tag{2.15}$$

where $\hat{y}_j$ and $\bar{y}$ indicate the estimated value and the mean of dependent observations, respectively.

The SSE basically measures the variation of modeling errors. In general, a small value of the SSE indicates that the model of regression can better interpret the data, while a big value of the SSE indicates that the model interprets the data poorly. The SSE can

be found (see Figure 2.2) by

$$SSE = \sum_{j=1}^{n} (y_j - \hat{y}_j)^2 = \sum_{j=1}^{n} e_j^2. \tag{2.16}$$

where y_j and $\hat{y}_j$ indicate the dependent observations and the estimated value, respectively.

The total sum of squares can be decomposed into the sum of squares explained by the regression and the sum of square errors as seen in Figure 2.2 [67, 68, 69, 70] as follows,

$$SST = SSR + SSE. \tag{2.17}$$

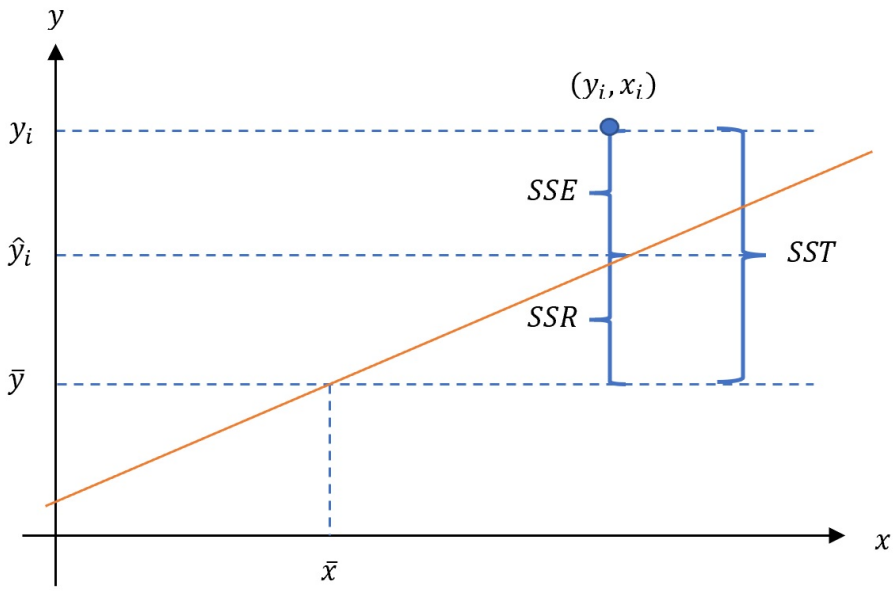


Figure 2.2 Explanation of the sum of squares

2.3.1.4 Determination of the Coefficient

Determination of the coefficient is a measure used in statistical analysis that assesses the model success in interpreting and predicting future results. It indicates the level of variance shown in the dataset. Determination of the coefficient, also known as R^2, is used as a guideline to measure the accuracy of the model,

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}. \tag{2.18}$$

It is the square of the correlation coefficient. The goodness-of-fit of the regression

line in the estimation of the dependent variable uses the independent variable. In other words, R^2 is a measure showing the rate of the contribution of the independent variables in the change of dependent variable. It ranges between zero to one, $0 \leq R^2 \leq 1$. When R^2 tends to be very high and closer to 1, the relationship is better, and a model becomes very reliable for future prediction. However, small R^2 does not imply that the model is bad. On the other hand, a value 0 indicates that the model fails to accurately design the data. It also allows R^2 to display the degree of correlation between the variables of interest [67, 68, 69, 70].

2.3.1.5 Residual Analysis

The residual of the observed value is the difference between the observed value and the estimated value. In regression analysis, the observations y_i may be different from the fitted values $\hat{y}_i$ (the predicted value) obtained from the model (see Figure 2.3). The vector of residuals, e_i, is thus given by:

$$e_i = y_i - \hat{y}_i. \tag{2.19}$$

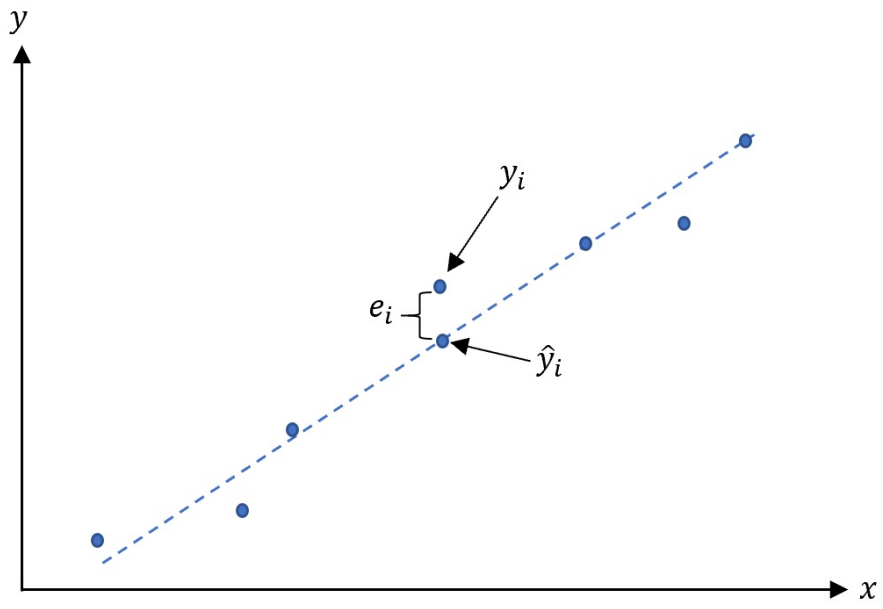


Figure 2.3 Explaining the residual

The mean square error (MSE) is the measure of the average square difference between the estimated values and the actual value. Also, the MSE of a regression is a number

computed from the sum of squares of the computed residuals. The smaller MSE closes to fit the data. Then it is given by

$$MSE = \frac{1}{n}\sum_{j=1}^{n} e_j^2. \tag{2.20}$$

The root mean square error (RMSE) is a common measure of the differences between sample values predicted by a model estimated and values observed [67, 68, 69, 70]. Then it is given by

$$RMSE = \sqrt{MSE}. \tag{2.21}$$

2.3.2 Particle Swarm Optimization

The PSO proposed by Kennedy and Earhart [72] has been used to solve various optimization problems. They inspired from social behavior of bird flocking or fish schooling, these animals have a major role in the development of the algorithm. So, the researchers used the PSO to estimate the parameters of models and implemented different strategies of mathematical methods to predict and to optimize problems. For instance, the PSO algorithm is applied to 28 well-known nonlinear regression models and the results display that the PSO algorithm provides accurate outcomes for estimating the parameter of their nonlinear regression models [73]. The PSO algorithm applied for finding the nonlinear model parameters [74], estimating the parameters of multiple linear regression models [75], the researchers used the PSO, genetic algorithm, and multiple regression in the estimation of soil mechanical resistance value [76], and estimation of the parameters was done for the nonlinear multi-regression model based on Choquet integral through a PSO algorithm [77].

The method optimizes a problem by trying to improve a solution. Each particle traces its coordinates in the area of problem that relates to the best solutions carried out so far. This value is called *Pbest*. Another "best" value that is tracked by the PSO is the best value, obtained so far by any particle in the neighbors of the particle. This location is called *lbest*. When the particle considers the whole population as its topological neighbors, the best value is a global best and is called *Gbest*. The PSO idea consists of, at each time step, changing the velocity of each particle towards the *Pbest* and *lbest* locations.

In the PSO, simple software agents, called particles, move in the search space for improvement. These randomly selected particles search solution space using the information of their neighborhood, personal information, and randomness. The position of a particle represents a candidate solution to the existing improvement problem. All particles look for better sites in the search space by changing their velocity at the end of each iteration. Because of each iteration, the position and velocity vectors are expressed as follows:

$$V_i^{t+1} = \omega V_i^t + c_1 r_1 (P_{best} - X_i^t) + c_2 r_2 (G_{best} - X_i^t) \tag{2.22}$$

$$X_i^{t+1} = X_i^t + V_i^{t+1} \tag{2.23}$$

where $t, \omega, c_1, c_2, r_1, r_2, V_i^t, X_i^t, P_{best}$ and G_{best} indicate iteration number, weight parameter, acceleration coefficients (cognitive parameter, social parameter), random numbers uniformly distributed between 0 and 1, velocity of individual i at iteration t, position of individual i at iteration t, the best local value of each particle, the best value of swarm, respectively [78, 79, 80].

We can see how the best position of the particle, *Pbest*, and the best position of the group, *Gbest*, affect the velocity of the particle in the next iteration. Therefore, the essential concept of the PSO is to accelerate each particle to the position of *Pbest* and *Gbest*, with a random weighted acceleration at each step (see Figure 2.4).

The update velocity for particles consists of three components in equations (2.22) and (2.23), in the two-dimensional search space. Therefore, Figure 2.4, illustrates how the three components of particle velocity move to the best global position in time steps *t* and *t+1*, respectively [81].

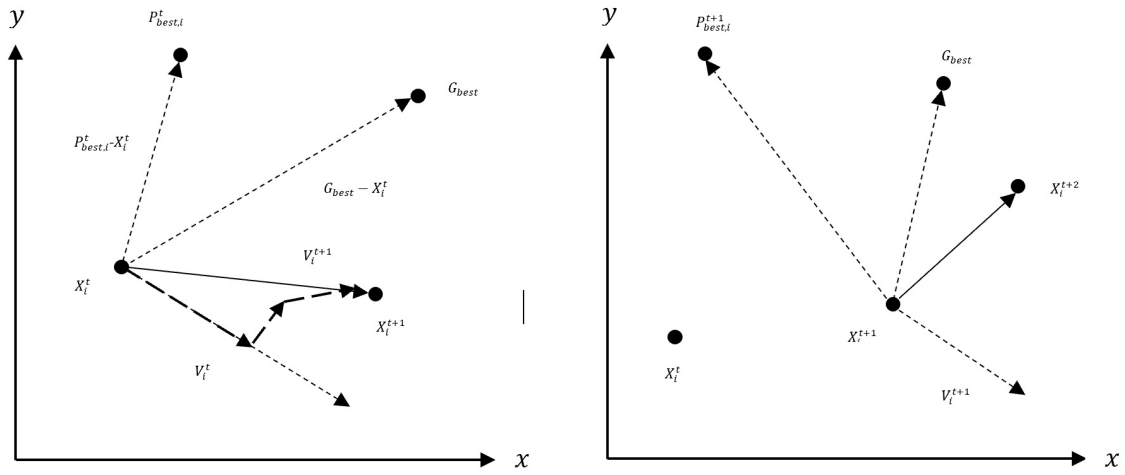


Figure 2.4 Update of a velocity and position for a particle in a 2D search space

2.4 Why These Methods

A major advantage of data analysis using the multiple regression model is the ability to determine the relative effect of one or more predictor variables on the value. Also, the estimates from a wide category of potential parameter estimates are used under the usual assumptions for the process modeling. Moreover, it uses the data very efficiently, and good results can be available with small data sets. Also, as an extremely important feature in the regression, the ideal parameters are obtained from the least squares regression by evaluating unknown parameters.

Despite the recognized advantages of conventional methods, most of them suffer from various disadvantages such as high cost, difficulty in use, and time-consuming. In this case, optimization can be recalled as a very good alternative to the corresponding methods. In the recent years, the PSO has been successfully applied to many areas to simplify optimization problems that had previously experienced serious difficulties. It is demonstrated that the PSO produces better results in a faster, cheaper way and the simplicity of the implementation, is the most attractive feature of this algorithm. Another reason that makes the PSO attractive is that it is reliable, robust, and considered as an effective meta-heuristic optimization algorithm.

3

ANEMIA MODELLING USING THE MULTIPLE LINEAR REGRESSION ANALYSIS

3.1 Introduction

A mathematical model is an essential tool for analyzing pathological characteristics and it can be used for various reasons as in the literature [10, 11, 12, 82, 83, 84, 85, 86, 87, 88]. To assess situations seen in hospitals, any disease condition has several effects for a single disease. So, most outcomes in real life problems are affected by multiple input variables.

This chapter aims at predicting pathological subjects from a population through physical biomedical variables (eight blood variables, sex, and age) and output (Anemia types). It is important to predict the type of anemia because there has been an increase in the incidence of anemia among different segments of society. To make the best biomedical decisions, medical predictions play a very important role in the process of diagnosis and planning treatment for health providers. So, our goal is to develop a new mathematical model to study the effect of the blood variables, sex, and age on the types of anemia. Our model, different from the mathematical models given in the literature [38, 89, 47, 90, 48] has also been successfully used in the prediction of several types of anemia through a large group of blood variables, sex, and age.

To the best knowledge of the author, more general models representing the behaviour closer to nature have been produced for the first time. The more number of input variables makes the derived model more realistic in the biomedicine. Thus, for such a realistic model, for such a large number of input variables a study has been accomplished here. Therefore, this study is believed to be an important contribution to predict the types of anemia.

Despite very effective, striking and frontier studies in the literature, researchers have used models with limited number of variables. Therefore, the present study focuses on the determination of the type of anemia through a very large number of the observational variables, more realistic one. Since many researchers have commonly considered the MRA among the modelling techniques to deal with various problems including anemia [38, 60, 61, 63, 91, 92, 93, 94, 95, 96, 97], the multiple analysis is taken into account in modelling the current biomedical problem.
The remainder of the chapter is organized as follows: Section 3.2 highlight the study samples, explain linear regression analysis procedure and test the model. Building the linear model of data by the regression analysis has been given in Section 3.3. Regression model has been analyzed and discussed in Section 3.4. Finally, conclusions and future research directions have been detailed.

3.2 Materials and Methods

3.2.1 Study Samples

As pointed out by the corresponding researchers, anemia is one of the most common blood diseases worldwide. The diagnosis of anemia depends on the concentration of hemoglobin less than the normal limits followed by the World Health Organization (WHO), and it is worth noting that the concentration of hemoglobin varies by age and sex as seen in Table 3.1 [43].
Anemia is classified into several types and those types differ in terms of their causes. Some types of anemia are hereditary. These types may affect children and may cause health problems for a lifetime. Women after adulthood may experience iron deficiency anemia, blood loss during the menstrual cycle, the most common type, may occur during pregnancy due to excessive need of minerals in the blood by the fetus during pregnancy, older people may be exposed to anemia due to malnutrition and other medical conditions [43].

Table 3.1 Hemoglobin thresholds used to define anemia [43]

Age or gender group	Hemoglobin threshold (g/l)
Children (0.50–4.99 yrs)	110
Children (5.00–11.99 yrs)	115
Children (12.00–14.99 yrs)	120
Non-pregnant women (≥15.00 yrs)	120
Pregnant women	110
Men (≥15.00 yrs)	130

The data were collected from observations of blood variables in order to identify a healthy or infected person and involved 539 subjects provided from blood laboratories in Iraq. Individuals between 6-56 years old have been taken into consideration and included 248 males, 291 females. Subjects are consisting of 211 healthy ones and of 328 anemic ones to build the model. The number of variables studied and selected for building the model is eleven, the independent variables identified are ten and a dependent variable. The dependent variable consists of six different outputs are healthy (0) and five blood diseases are iron deficiency anemia (1), deficiency vitamin B12 (2), thalassemia (3), sickle cell (4) and spherocytosis (5).

Here the samples for people and for each subject readings of blood variables are [43, 44] Hemoglobin (HB), Red Blood Cells (RBC), Mean Corpuscular Hemoglobin (MCH), White Blood Cell (WBC), Hematocrit (HCT), Mean Corpuscular Hemoglobin Concentration (MCHC), Platelets (PLT), Mean Corpuscular Volume (MCV) and sex and age. The anemia types and blood variables for our data are displayed in Table 3.2.

Table 3.2 Some samples from the data

HB	RBC	MCH	WBC	MCV	HCT	MCHC	PLT	Sex	Age	Anemia type
17.5	5.55	31.6	14.1	92	50.9	34.5	318	2	23	0
16.3	6.07	26.9	8.16	80.9	49.1	33.2	349	1	23	0
11.1	4.38	25.3	5.8	81	35.6	31.1	227	1	11	1
11.1	4.85	22.8	10	81	39.4	28.1	274	2	16	1
9	3.47	25.8	2.3	88	30.4	29.5	148	1	11	2
1.46	4.4	30.4	59.8	108	15.8	28.2	330	2	29	2
8.1	3.6	22.4	12	78	28.1	28.7	472	1	15	3
3.92	6.6	16.8	8.3	60	23.7	27.9	443	2	17	3
8.3	2.58	31.9	12.4	103	26.7	30.9	458	1	11	4
7.9	2.88	27.4	17.55	83	23.9	33.1	703	1	16	4
6.8	5.77	11.7	11.9	49	28.4	23.8	573	2	11	5

3.2.2 Multiple Linear Regression Model

Consider a MLR model with k predictor variables, independent observations

$$\mathbf{y} = B_0 + B_1 x_1 + B_2 x_2 + ... + B_k x_k + \epsilon = B_0 + \sum_{i=1}^{k} B_i x_i + \epsilon. \tag{3.1}$$

The observations recorded for each of these n levels can be expressed in the following way

$$\begin{aligned}
y_1 &= B_0 + B_1 x_{11} + B_2 x_{12} + ... + B_k x_{1k} + \epsilon_1 \\
y_2 &= B_0 + B_1 x_{21} + B_2 x_{22} + ... + B_k x_{2k} + \epsilon_2 \\
&\vdots \\
y_i &= B_0 + B_1 x_{i1} + B_2 x_{i2} + ... + B_k x_{ik} + \epsilon_i \\
&\vdots \\
y_n &= B_0 + B_1 x_{n1} + B_2 x_{n2} + ... + B_k x_{nk} + \epsilon_n
\end{aligned} \tag{3.2}$$

The dependent observations $y_1, y_2, ..., y_n$, and the independent observations x_1, x_2 $, ..., x_k$, have n levels. Then x_{ij} represents the ith level of the jth predictor variable, x_j.

System (3.2) can be represented as follows:

$$\mathbf{y} = \mathbf{BX} + \epsilon, \tag{3.3}$$

with

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix}, \mathbf{B} = \begin{bmatrix} B_0 \\ B_1 \\ \vdots \\ B_k \end{bmatrix}, \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \tag{3.4}$$

where **y**,**X**,**B** and ϵ stand for the observations, the regression coefficients and an unobserved random variable that adds noise to the linear relationship between the dependent variable and regressors, respectively.
To obtain the regression model, **B** should be known. Therefore, **B** is estimated by using the least square estimates as follows

$$\hat{\mathbf{B}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}, \tag{3.5}$$

where $\mathbf{X}^T$ represents the transpose of the matrix **X** while $(\mathbf{X}^T\mathbf{X})^{-1}$ represents inverse of the matrix $(\mathbf{X}^T\mathbf{X})$. Knowing the estimate $\hat{\mathbf{B}}$, the MLR model can now be expressed as [66, 67]

$$\hat{\mathbf{y}} = \hat{\mathbf{B}}\mathbf{X}, \tag{3.6}$$

where $\hat{\mathbf{y}}$ is the estimated value for **y** from the regression.

3.2.3 Test for the Model

The linear regression model estimation is selected and the sum of square tests. The computation formula can be given as follows:

$$SST = \sum_{j=1}^{n}(y_j - \bar{y})^2, \tag{3.7}$$

$$SSR = \sum_{j=1}^{n}(\hat{y}_j - \bar{y})^2, \tag{3.8}$$

$$SSE = \sum_{j=1}^{n} (y_j - \hat{y}_j)^2 = \sum_{j=1}^{n} e_j^2. \tag{3.9}$$

The coefficient of determination is a measure showing the rate of the contribution of the independent variables in the interpretation of the change in the dependent variable as known from the literature [70, 71]. It is given as follow:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}. \tag{3.10}$$

A terminological difference arises in the expression mean squared error (MSE). The MSE of a regression is a measure of the average of the sum of squared error and how the concentration of data around the regression model. The smaller the MSE, whenever the results are more accurate [70, 71]. Then it is given by

$$MSE = \frac{1}{n} \sum_{j=1}^{n} e_j^2. \tag{3.11}$$

3.3 Building Linear Regression Analysis Model

The currently produced MLR model is a linear equation determined as previously mentioned in Section 3.2.2. The obtained model is as follows:

$$\begin{aligned} \mathbf{y} =& B_0 + B_1 HB + B_2 RBC + B_3 MCH + B_4 WBC + B_5 MCV \\ &+ B_6 HCT + B_7 MCHC + B_8 PLT + B_9 Sex + B_{10} Age + \epsilon \end{aligned} \tag{3.12}$$

where $\mathbf{y}$ is type of the anemia and B_i, $0 \leq i \leq 10$, are the parameters to be determined. The linear regression model, as explained in Section 3.2.2, is estimated as

$$\begin{aligned} \hat{\mathbf{y}} =& 6.377 - 0.224 HB - 0.224 RBC - 0.029 MCH + 0.001 WBC \\ &+ 0.0005 MCV - 0.016 HCT + 0.007 MCHC + 0.001 PLT \\ &- 0.311 Sex - 0.009 Age. \end{aligned} \tag{3.13}$$

Here the coefficient values of the linear model have been obtained through the multiple regression approach, to find the model that is more realistic (see Table 3.9).

As previously mentioned, the model can be represented in matrix form as follows:

$$\hat{\mathbf{y}} = \hat{\mathbf{B}}\mathbf{X} \tag{3.14}$$

where

$$\hat{\mathbf{y}} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{539} \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & HB_{11} & RBC_{12} & \dots & Age_{110} \\ 1 & Hb_{21} & RBC_{22} & \dots & Age_{210} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & HB_{539,1} & RBC_{539,2} & \dots & Age_{539,10} \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 6.377 \\ -0.224 \\ -0.224 \\ -0.029 \\ 0.001 \\ 0.0005 \\ -0.016 \\ 0.007 \\ 0.001 \\ -0.311 \\ -0.009 \end{bmatrix} \tag{3.15}$$

Here $\hat{\mathbf{y}}$ and $\mathbf{X}$ represent the estimates for output (types of the anemia) and the independent observations matrix, respectively.

3.4 Results and Discussion

Different strategies of mathematical methods are implemented to analyze blood variables, as in the literature [47, 48, 50, 98]. The MRA has been taken into account by many researchers [38, 60, 61, 63, 91, 92, 93, 94, 95, 96, 97] while dealing with various anemia problems at different levels. However, they used a limited number of blood variables and they did not study a relationship for the prediction of the types of anemia. Therefore, the current study concentrates on the investigation of the relationship between a very large number of blood variables and the types of anemia. Various versions of models, based on the variables, are derived (see Table 3.3).

Table 3.3 Various forms of the multiple linear models: blood variables, sex, and age.

Models	R^2	MSE
Model 1 for (HB, sex and age)	0.568	0.935
Model 2 for (RBC, sex and age)	0.174	1.787
Model 3 for (MCH, sex and age)	0.255	1.611
Model 4 for (WBC, sex and age)	0.229	1.667
Model 5 for (MCV, sex and age)	0.190	1.752
Model 6 for (HCT, sex and age)	0.649	0.759
Model 7 for (MCHC, sex and age)	0.243	1.637
Model 8 for (PLT, sex and age)	0.271	1.577
Model 9 for (HB, RBC, sex and age)	0.686	0.680
Model 10 for (MCH, WBC, sex and age)	0.304	1.509
Model 11 for (MCV, HCT, sex and age)	0.649	0.760
Model 12 for (MCHC, PLT, sex and age)	0.314	1.486
Model 13 for (WBC, MCV, HCT, MCHC, sex and age)	0.668	0.723
Model 14 for (HB, RBC, MCH, PLT, sex and age)	0.698	0.656

The models produced in terms of larger number of blood variables show better correlation than the models produced in terms of less number of blood variables for predicting the types of anemia in equation (3.13). However, naturally some of the variables are of more effect than others.

After the essential requirements have been verified for the multivariate analysis in equation (3.13), the variables have been included for the MLR analysis. Those variables consist of regression coefficients B, the blood variables (HB, RBC, MCH, WBC, MCV, HCT, MCHC, PLT), sex, and age. Therefore, the MLR shows the synergistic effect of predicting the types of anemia better than the ones used fewer blood variables. The enter method of the MLR has been used in the current analysis. All the variables were introduced into the regression model as selected by the enter method of the MLR.

In the outcome of the current analysis, it has been found that there is a more significant relation (R^2=0.699) of the MLR model. It means that 69.90% of the change in the relationship between all blood variables, sex, and age for the types of anemia is

explained.

Also, the diagnosis of anemia depends on hemoglobin thresholds used to define anemia followed by the WHO for age, and it is worth noting that the concentration of hemoglobin varies by age. Here we classify the data into three categories as age (6-11) years old, (12-14) years old, and ($\geq$15) years old as seen in Table 3.1 [43].

We have compared the results for the age group (6-56) with other classified age groups (6-11), (12-14), and (15-56). It has been found out that the results produced for the age group (6-56) are better than all other classified groups (see Tables 3.4-3.7). This difference is believed to stem from the decreasing the data as seen in Table 3.5.

In the outcome of the current analysis, it has been found that there is more significant relation of the MLR model for the data (6-56) comparison to the other cases (6-11), (12-14), and (15-56). It explains 69.90% of the change in the relationship between all blood variables, sex, age and the types of anemia as seen in Table 3.7. It is the best comparison to the results 48.2%, 83.8%, and 68.6% for the three categories (6-11), (12-14), and (15-56), respectively, as seen in Tables 3.4-3.6.

Table 3.4 Various forms of linear regression models: Blood variables, sex and age (6-11).

Models	R^2	RMSE
Model 1 for (HB, sex and age)	0.315	0.82956
Model 2 for (RBC, sex and age)	0.247	0.86971
Model 3 for (WBC, sex and age)	0.034	0.98506
Model 4 for (PLT, sex and age)	0.045	0.97960
Model 5 for (HB, RBC, sex and age)	0.333	0.82221
Model 6 for (MCH, WBC, sex and age)	0.039	0.98658
Model 7 for (MCV, HCT, sex and age)	0.385	0.78956
Model 8 for (MCHC, PLT, sex and age)	0.092	0.95889
Model 9 for (HB, RBC, MCH, sex and age)	0.375	0.79909
Model 10 for (WBC, MCV, HCT, sex and age)	0.417	0.77142
Model 11 for (HB, MCHC, PLT, sex and age)	0.392	0.78824
Model 12 for (WBC, MCV, HCT, MCHC, sex and age)	0.420	0.77270
Model 13 for (HB, RBC, MCH, PLT, sex and age)	0.431	0.76579
Model 14 for (HB, RBC, MCH, WBC, MCV, HCT, MCHC, PLT, sex and age)	0.482	0.74322

Table 3.5 Various forms of linear regression models: Blood variables, sex and age (12-14).

Models	R^2	RMSE
Model 1 for (HB, sex and age)	0.666	0.64123
Model 2 for (RBC, sex and age)	0.567	0.73044
Model 3 for (WBC, sex and age)	0.201	0.99263
Model 4 for (PLT, sex and age)	0.261	0.95426
Model 5 for (HB, RBC, sex and age)	0.678	0.64979
Model 6 for (MCH, WBC, sex and age)	0.383	0.89857
Model 7 for (MCV, HCT, sex and age)	0.385	0.89766
Model 8 for (MCHC, PLT, sex and age)	0.536	0.77954
Model 9 for (HB, RBC, MCH, sex and age)	0.755	0.58501
Model 10 for (WBC, MCV, HCT, sex and age)	0.397	0.91793
Model 11 for (HB, MCHC, PLT, sex and age)	0.798	0.53143
Model 12 for (WBC, MCV, HCT, MCHC, sex and age)	0.580	0.79267
Model 13 for (HB, RBC, MCH, PLT, sex and age)	0.757	0.60328
Model 14 for (HB, RBC, MCH, WBC, MCV, HCT, MCHC, PLT, sex and age)	0.838	0.58174

Table 3.6 Various forms of linear regression models: Blood variables, sex and age (15-56).

Models	R^2	RMSE
Model 1 for (HB, sex and age)	0.562	0.95026
Model 2 for (RBC, sex and age)	0.034	1.41107
Model 3 for (WBC, sex and age)	0.134	1.33618
Model 4 for (PLT, sex and age)	0.193	1.28994
Model 5 for (HB, RBC, sex and age)	0.677	0.81736
Model 6 for (MCH, WBC, sex and age)	0.272	1.22649
Model 7 for (MCV, HCT, sex and age)	0.638	0.86461
Model 8 for (MCHC, PLT, sex and age)	0.295	1.20694
Model 9 for (HB, RBC, MCH, sex and age)	0.680	0.81427
Model 10 for (WBC, MCV, HCT, sex and age)	0.642	0.86102
Model 11 for (HB, MCHC, PLT, sex and age)	0.580	0.93290
Model 12 for (WBC, MCV, HCT, MCHC, sex and age)	0.665	0.83370
Model 13 for (HB, RBC, MCH, PLT, sex and age)	0.685	0.80897
Model 14 for (HB, RBC, MCH, WBC, MCV, HCT, MCHC, PLT, sex and age)	0.686	0.81159

Table 3.7 Various forms of linear regression models: Blood variables, sex and age (6-56).

Models	R^2	RMSE
Model 1 for (HB, sex and age)	0.568	0.96685
Model 2 for (RBC, sex and age)	0.174	1.33677
Model 3 for (WBC, sex and age)	0.229	1.29106
Model 4 for (PLT, sex and age)	0.271	1.25593
Model 5 for (HB, RBC, sex and age)	0.686	0.82466
Model 6 for (MCH, WBC, sex and age)	0.304	1.22844
Model 7 for (MCV, HCT, sex and age)	0.649	0.87201
Model 8 for (MCHC, PLT, sex and age)	0.314	1.21903
Model 9 for (HB, RBC, MCH, sex and age)	0.692	0.81825
Model 10 for (WBC, MCV, HCT, sex and age)	0.656	0.86483
Model 11 for (HB, MCHC, PLT, sex and age)	0.582	0.95278
Model 12 for (WBC, MCV, HCT, MCHC, sex and age)	0.668	0.85008
Model 13 for (HB, RBC, MCH, PLT, sex and age)	0.698	0.80985
Model 14 for (HB, RBC, MCH, WBC, MCV, HCT, MCHC, PLT, sex and age)	0.699	0.81171

Thus, it is concluded that the regression model with the blood variables, sex, and age are seen to be significant ($p < 0.000$). That means simultaneous consideration of the blood variables, sex, and age has a significant effect on the relationship on the determination of the types of anemia (see Table 3.8).

Table 3.8 Analysis of the variance for the correlation in equation (3.13)

	Sum of Squares	Degrees of freedom	Mean Square	F-Stat	P-Value
Regression	809.354	10	80.935	122.838	0.000
Residual	347.889	528	0.659		
Total	1157.243	538			

The standardized coefficient (Beta) compares the effect force of each individual blood variables, sex, and age to the types of anemia. It is thus given by $StandardizedBeta_j = B_j * SD(X_j)/SD(Y)$.

The HB absolute value of the Beta coefficient is (-0.663) has the strongest relationship

with the types of the disease comparison to the other variables RBC (−0.345), Sex (−0.106), HCT (−0.100), MCH (−0.090), PLT (0.080), Age (−0.065), WBC (0.016), MCHC (0.016) and MCV (−0.001). The interpretation of the Beta value for the HB signifies that for every change in the HB, the dependent variable will be changed by the Beta coefficient value (see Table 3.9).

The *t*-test was used to measure the partial effect of the variables HB, RBC, MCH, WBC, MCV, HCT, MCHC, PLT, sex, and age on the types of anemia. Notice that these variables have been seen to affect the types of anemia but in varying rates (see Table 3.9). The histogram of the residuals which confirm that the data are distributed according to a normal distribution with a mean of zero and a standard deviation of 0.991 (see Figure 3.1).

Table 3.9 Analysis of the multiple regression coefficients given in equation (3.13)

	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	*t*-Stat	*P*-Value
(Const.)	6.377	0.552		11.563	0.000
HB	-0.224	0.062	-0.663	-3.581	0.000
RBC	-0.224	0.066	-0.345	-3.392	0.001
MCH	-0.029	0.015	-0.090	-1.931	0.054
WBC	0.001	0.003	0.016	0.549	0.583
MCV	0.0005	0.008	-0.001	-0.015	0.988
HCT	-0.016	0.028	-0.100	-0.575	0.565
MCHC	0.007	0.016	0.016	0.464	0.643
PLT	0.001	0.000	0.080	2.637	0.009
Sex	-0.311	0.074	-0.106	-4.191	0.000
Age	-0.009	0.004	-0.065	-2.303	0.022

To find out the extent of spread the random error around the linear regression model, the MLR use the mean square residuals, MSE=0.659 (see Table 3.8). Small values of the MSE indicate the concentration of data around the linear regression model (see Figure 3.2).

Table 3.10 Comparison of the MLR results with the results of the linear deep learning method

Methods	SSE	MSE	R^2
Linear Regression Analysis	347.889	0.659	0.699
Linear Deep Learning Methods (LSTM)	349.869	0.665	0.695

LSTM: Long Short Term Memory

In this study, comparing criteria are constructed on the principle of whether the technique provides a suitable prediction or not. This task is achieved by comparing with the deep learning method (LSTM) [99]. The results demonstrate that the linear regression has the best fit to the initial dataset comparing to the deep learning method (LSTM) (see Table 3.10). Therefore, the present study provides an accurate model for prediction of the types of anemia.

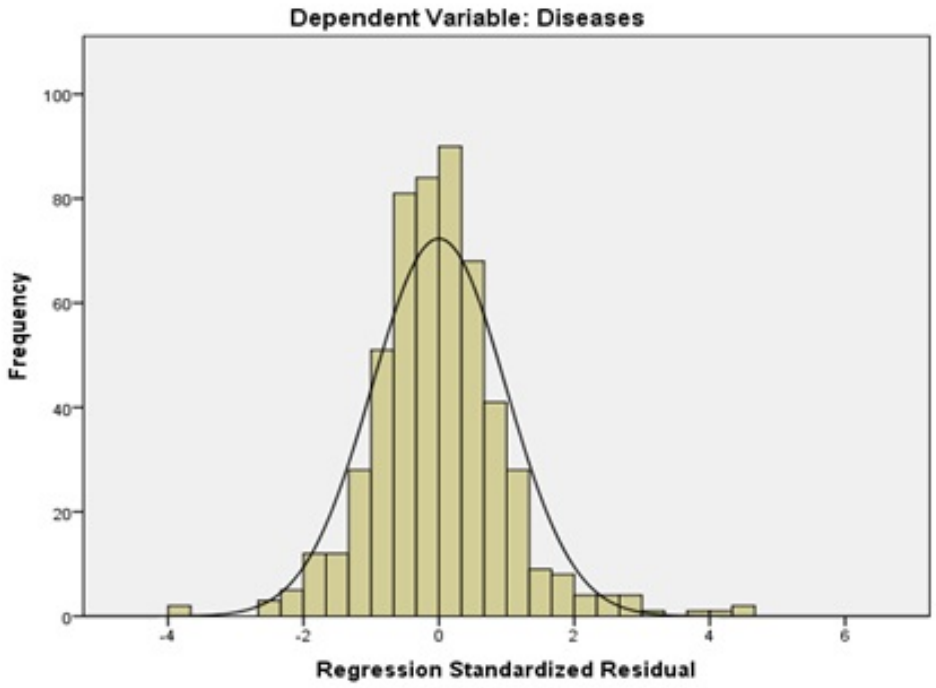


Figure 3.1 Histogram of the residuals

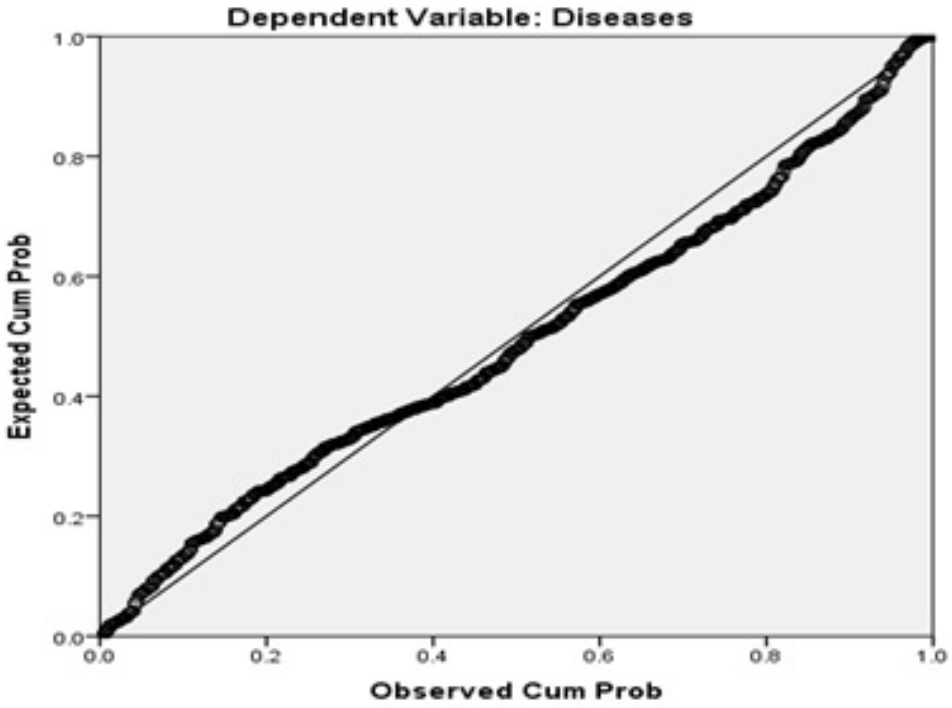


Figure 3.2 Normal P–P Plot of Regression Standardized Residual

3.5 Conclusions

The MLR model, for the first time, have been derived in forecasting the types of anemia. The results revealed that the regression model is very promising and is capable of making the prediction. In the analysis of the current anemia problem, the multiple regression method has been found to be more accurate than linear deep learning methods. It has been concluded that the model is expected to be helpful for diagnosis of the types of anemia to health providers and designing appropriate treatment programs for their patients.

4

ANEMIA PREDICTION WITH MULTIPLE NONLINEAR REGRESSION ANALYSIS

4.1 Introduction

A mathematical model is a platform for understanding the behavior of a physical or a biophysical system. Mathematical modelling can be used for various reasons. How well any particular objective achieved depends on both the state of knowledge about the system and how well the modelling is done. As seen in the literature [10, 12, 83, 84, 87, 100, 101], mathematical modeling has been shown to be an essential tool for also analyzing pathological characteristics. To assess situations seen in hospitals, any disease condition has several effects (inputs) for a single disease (output). So, most outcomes in real life problems are affected by multiple input variables. To understand such relationships, the used models that consider more than one input to produce a single output. As signified in the literature [45, 46, 102], the anemia of chronic inflammation and it was initially thought to be associated primarily with the infectious, inflammatory diseases.

This chapter aims at predicting pathological subjects from a population through physical observational variables (eight blood variables, sex, and age) and output (types of disease). It is important to predict the type of anemia because there has been an increase in the incidence of anemia among different segments of society. To make the best biomedical decisions, medical predictions play a very important role in the process of diagnosis and planning treatment for health providers. Thus, our goal is to derive a new mathematical model to study the effect of the blood variables, sex, and age on the types of anemia. Our model, differ from the mathematical models given in the literature [37, 50, 89, 47, 90, 48], have also been successfully used in the pre-

diction of several types of anemia through a large group of blood variables, sex, and age.

To the best knowledge of the authors, a more general model representing the behaviour closer to nature have been produced for the first time. The more number of input variables makes the derived model more realistic in the biomedicine. Thus, for such a realistic model, for such a large number of input variables, a study has been accomplished here. Therefore, this study is believed to be an important contribution to predict the types of anemia.

Despite very effective, striking and frontier studies in the literature, researchers have used models with a limited number of variables. Therefore, the present study focuses on the determination of the type of anemia through a very large number of the observational variables, more realistic one. Since many researchers have commonly considered the MRA among the modelling techniques to deal with various problems including anemia [37, 38, 60, 61, 62, 63, 64, 65, **?**, 92, 93, 94, 95, 97, 103], the multiple analysis is taken into account in modelling the current biomedical problem based on estimating optimum values in the set of the fitting parameters of the model.

The remainder of the chapter is organized as follows: Section 4.2 highlight the study samples, explain nonlinear regression analysis procedure and test the model. Building the model of data by the regression analysis has been given in Section 4.3. The produced results for the model are given in Section 4.4. The regression model has been analyzed and discussed in Section 4.5. Finally, conclusions and future research directions have been detailed.

4.2 Materials and Methods

4.2.1 Study Samples

Here the samples for people and for each subject readings of blood variables are Hemoglobin (HB), Red Blood Cells (RBC), Mean Corpuscular Hemoglobin (MCH), White Blood Cell (WBC), Hematocrit (HCT), Mean Corpuscular Hemoglobin Concentration (MCHC), Platelets (PLT), Mean Corpuscular Volume (MCV) and sex and age [43, 44].

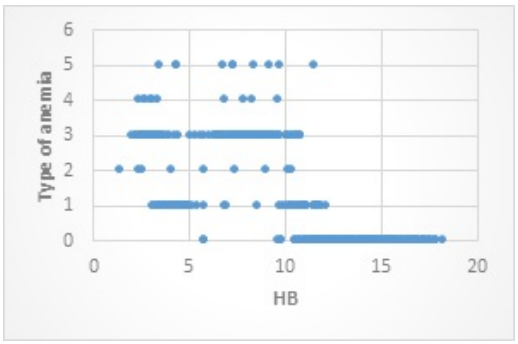
For the data, it is considered that blood diseases are iron deficiency anemia (1), de-

ficiency vitamin B12 (2), thalassemia (3), sickle cell (4) and spherocytosis (5). The anemia types and blood variables for our data are displayed in Table 4.1 and Figure 4.1.

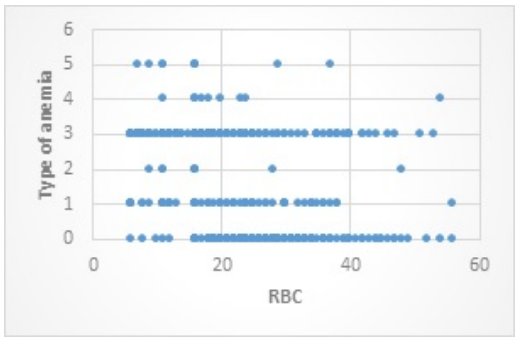
Table 4.1 Some samples from the data

HB	RBC	MCH	WBC	MCV	HCT	MCHC	PLT	Sex	Age	Anemia type
17.5	5.55	31.6	14.1	92	50.9	34.5	318	2	23	0
16.3	6.07	26.9	8.16	80.9	49.1	33.2	349	1	23	0
11.1	4.38	25.3	5.8	81	35.6	31.1	227	1	11	1
11.1	4.85	22.8	10	81	39.4	28.1	274	2	16	1
9	3.47	25.8	2.3	88	30.4	29.5	148	1	11	2
1.46	4.4	30.4	59.8	108	15.8	28.2	330	2	29	2
8.1	3.6	22.4	12	78	28.1	28.7	472	1	15	3
3.92	6.6	16.8	8.3	60	23.7	27.9	443	2	17	3
8.3	2.58	31.9	12.4	103	26.7	30.9	458	1	11	4
7.9	2.88	27.4	17.55	83	23.9	33.1	703	1	16	4
6.8	5.77	11.7	11.9	49	28.4	23.8	573	2	11	5

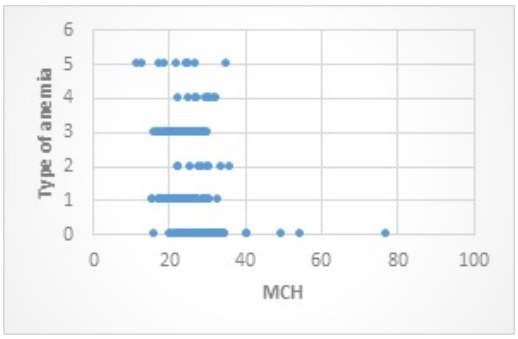
The chapter aims at predicting pathological subjects from a population in terms of various biomedical information. Therefore, the data were collected from observations of blood variables in order to identify a healthy or infected person and involved 539 subjects provided from blood laboratories in Iraq. Individuals between 6-56 years old have been taken into consideration and included 248 males, 291 females. Subjects are consisting of 211 healthy ones and of 328 anemic ones to build the model. The dependent variable consists of six different outputs (healthy: 0 and five blood diseases: 1-5). Therefore, the corresponding dependent and independent variables based on data are used to improve health standards of individuals.


Type of anemia
HB
0
5
10
15
20

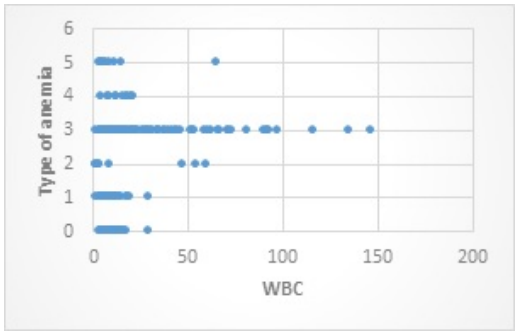
(a)


Type of anemia
RBC
0
20
40
60

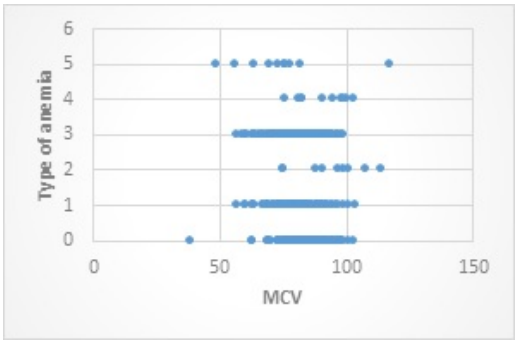
(b)


Type of anemia
MCH
0
20
40
60
80
100

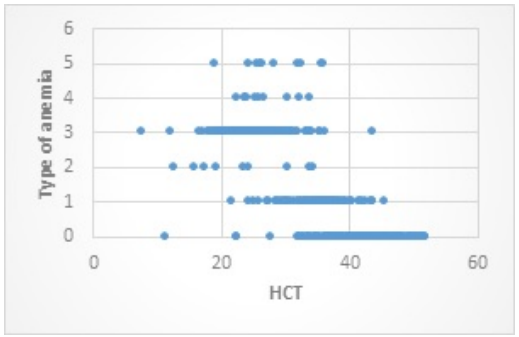
(c)


Type of anemia
WBC
0
50
100
150
200

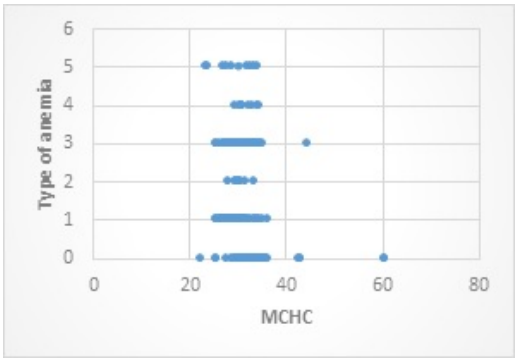
(d)


Type of anemia
MCV
0
50
100
150

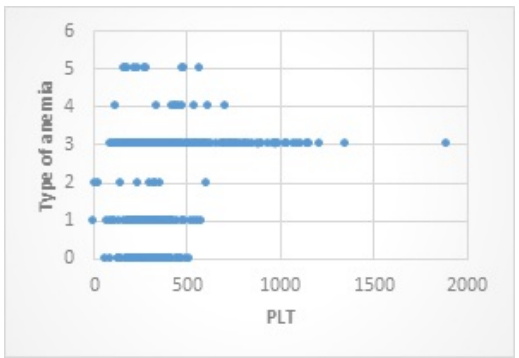
(e)


Type of anemia
HCT
0
20
40
60

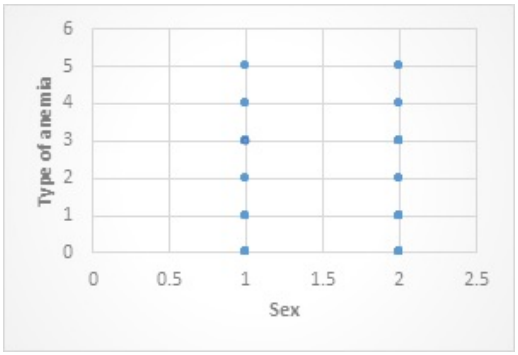
(f)


Type of anemia
MCHC
0
20
40
60
80

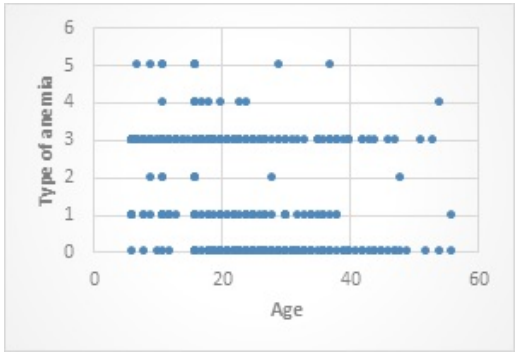
(g)


Type of anemia
PLT
0
500
1000
1500
2000

(h)



(i)



(j)

Figure 4.1 Anemia Types and blood variables: (a) HB and the anemia types; (b) RBC and the anemia types; (c) MCH and the anemia types; (d) WBC and the anemia types; (e) MCV and the anemia types; (f) HCT and the anemia types; (g) MCHC and the anemia types; (h) PLT and the anemia types; (i) Sex and the anemia types; (j) Age and the anemia types

4.2.1.1 Multiple Nonlinear Regression Model

Nonlinear regression analysis model can be given as

$$\mathbf{y} = f(\mathbf{X},\mathbf{B}) + \epsilon, \tag{4.1}$$

where $\mathbf{y}, \mathbf{X}, \mathbf{B}, f()$ and ϵ indicate the observations, the vector of the regression coefficients, the known nonlinear regression function and the unobserved random variable that adds noise to the nonlinear relationship between the dependent variable and regressors, respectively.

The basis of the method is to approximate the model by a linear one and to refine the parameters by successive iterations [68, 69]. First, let

$$y_i = f(X_i, B) + \epsilon_i, 1 \leq i \leq n, \tag{4.2}$$

and

$$Q = \sum_{i=1}^{k} (y_i - f(X_i, B))^2. \tag{4.3}$$

In order to find

$$\hat{B} = \arg\min_{B} Q, \tag{4.4}$$

first each of the partial derivatives of Q is found with respect to B_j. Then, each of

the partial derivatives is taken to be equal to 0 and the parameters B_k are replaced by $\hat{B}_k, 0 \leq k \leq n$. The functions to be solved are nonlinear in the parameter estimates $\hat{B}_k$. The regression analysis is using the optimization to estimate the parameters of the model by minimizing the sum of the squared error function. So, a form of optimization could be considered as the best way in selecting a suitable model [71, 104].

4.2.2 Test for the Model

The regression model estimation is selected with the confidence interval of 95% and adjusted sum of square tests (Type III). The computation formulae can be given as follows:

$$SST = \sum_{j=1}^{n} (y_j - \bar{y})^2, \tag{4.5}$$

$$SSR = \sum_{j=1}^{n} (\hat{y}_j - \bar{y})^2, \tag{4.6}$$

$$SSE = \sum_{j=1}^{n} (y_j - \hat{y}_j)^2 = \sum_{j=1}^{n} e_j^2. \tag{4.7}$$

The coefficient of determination is a measure showing the rate of the contribution of the independent variables in the interpretation of the change in the dependent variable as known from the literature [70, 71], small R^2 does not imply that the model is not significant. It is given as follow:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}. \tag{4.8}$$

4.2.3 Residual Analysis

In the regression analysis, the observations y_i may be different from the fitted values $\hat{y}_i$ obtained from the model. The difference between these two values is the residual, e_i. The vector of residuals, $\mathbf{e}_i$, is thus given by:

$$\mathbf{e}_i = y_i - \hat{y}_i. \tag{4.9}$$

A terminological difference arises in the expression mean squared error (MSE). The

MSE of a regression is a measure of the average of the sum of squared error and how the concentration of data around the regression model. The smaller the MSE, whenever the results are more accurate [70, 71, 104]. Then it is given by

$$MSE = \frac{1}{n}\sum_{j=1}^{n} e_j^2. \tag{4.10}$$

4.3 Building Nonlinear Regression Analysis Model

Important problems can usually be represented by mathematical models. Building multiple regression model of a data is one of the most challenging problems. Now, attention is paid to the model building process in the sense that it is attempted to find the best relation between the independent variables and the dependent variable **y** so that the final complete model is investigated in the regression model. Given the problem and data but without a model, the model building process can often be aided by graphs that help visualize the relationship between the different variables in data [70, 104]. Main steps in building a model of a dataset are given by conducting regression analysis (see Figure 4.2).

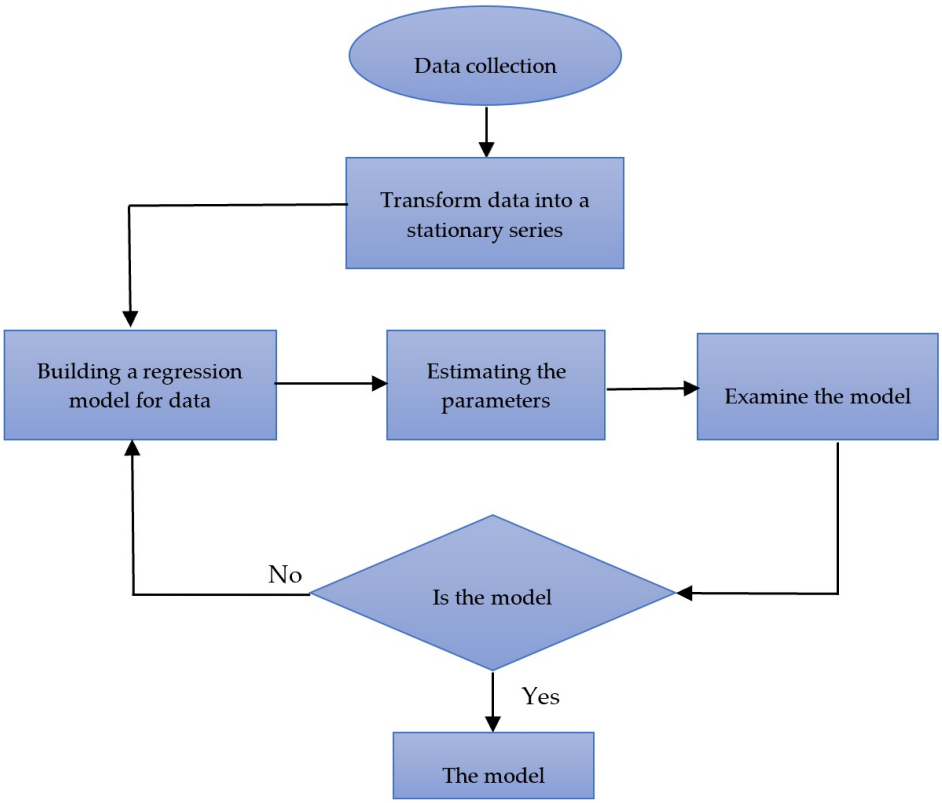


Figure 4.2 Main steps in the regression analysis procedure

In the regression model, linear regression is used first to determine whether a particular type of curve can be fit into our data. If enough fitness cannot be obtained by using the linear regression, then it may be needed to choose a nonlinear regression. Although the linear regression can represent curves, it is restricted in the forms of curves that can be contained for the data. Sometimes the curve specified in the data cannot be contained. Nonlinear regression can be suitable with many types of data, but it may require more effort to find the best fit and explain the role of independent variables.

For various approaches as pointed out in the literature [65, 98], a nonlinear model is usually expected to fit better than their linear rivals.

A nonlinear regression model describes a nonlinear relationship between the dependent and the independent variables. As is the case in the linear regression model, a multiple nonlinear model, based on the data, is built in Section 4.2.1.1. The produced model is as follows:

$$\hat{\mathbf{y}} = \frac{B_0}{E_1 + E_2} + \epsilon \tag{4.11}$$

$$E_1 = B_1(HB)^6 + B_2(RBC)^5 + B_3(MCH)^4 + B_4(WBC)^3 + B_5(Sex)^2$$
$$E_2 = B_6(HCT) + B_7(MCHC)^{\frac{1}{2}} + B_8(PLT)^{\frac{1}{3}} + B_9(MCV)^{\frac{1}{4}} + B_{10}(Age)$$

where $\mathbf{y}$ stands for the types of anemia. In principle, many nonlinear models can be proposed in dealing with the anemia problem. In this connection, here, several attempts have been made to obtain the best results, depending on the biomedical variables and the power of each input variable of interest. Other nonlinear model types for the data were also taken into consideration and it was observed that the current model was the best among these models to obtain accurate results. In addition, those powers of the variables in the model have been investigated and the regression analysis has been taken to find the optimum parameter values of the model (see Table 4.4), so as to obtain the best fitting for the data. It is well-known that the model order is chosen according to the number of bends you need on your structure. Each increase in the exponent produces one more bend in the fitted structure. Therefore, it is tried to be found the multivariable nonlinear function that best fits the specific structure in the data. The accepted nonlinear regression model is then estimated as

$$\hat{\mathbf{y}} = \frac{2489.986}{E_1 + E_2} \tag{4.12}$$

Here the denominator of equation is separated into two parts as E_1, E_2. Thus, the separated parts are clearly expressed as:
$E_1 = 0.001(HB)^6 + 0.014(RBC)^5 - 0.001(MCH)^4 + 0.0001(WBC)^3 + 18.711(Sex)^2$
$E_2 = -60.591(HCT) - 297.972(MCHC)^{\frac{1}{2}} - 26.450(PLT)^{\frac{1}{3}} + 1367.932(MCV)^{\frac{1}{4}} + 1.469(Age)$.
Here $\hat{\mathbf{y}}$ represent the estimates for the types of anemia. The coefficient values of the nonlinear model have been optimized for the multiple regression approaches, to find the more realistic model (see Table 4.4).

4.4 Nonlinear Model Results

This study here focuses on the discovery of a possible relationship between the blood variables and the types of anemia through the nonlinear model and explains the significance level of the model (see Table 4.2). Various types of models, based on different possibilities, have been produced through the biomedical variables (see Table 4.3).

The regression approach has been taken to forecast the best parameters with the optimum residual sum of squares value (see Table 4.4 and Figure 4.3). It is important to note that, the nonlinear deep learning (LSTM) and nonlinear regression neural network methods have also been used to compare our model results. The results revealed that the currently derived model is seen to be better than the other two rivals (see Table 4.5).

Table 4.2 Analysis of variance and R^2

Source	Sum of Squares	Degrees of freedom	Mean Squares
Regression	2185.852	11	198.714
Residual	271.148	528	0.514
Uncorrected Total	2457.000	539	
Corrected Total	1157.243	538	
R^2=1-(Residual Sum of Squares) /(Corrected Sum of Squares)	0.766		

Table 4.3 Various forms of the multiple nonlinear regression models: blood variables, sex, and age

Models	E_1, E_2 in model equation (4.12)	R^2	MSE
Model 1 for (HB, sex and age)	$E_1 = 0.001(HB)^6 + 18.711(Sex)^2$, $E_2 = 1.469(Age)$	0.551	0.971
Model 2 for (RBC, sex and age)	$E_1 = 0.014(RBC)^5 + 18.711(Sex)^2$, $E_2 = 1.469(Age)$	0.069	2.014
Model 3 for (MCH, sex and age)	$E_1 = -0.001(MCH)^4 + 18.711(Sex)^2$, $E_2 = 1.469(Age)$	0.000	4.490
Model 4 for (WBC, sex and age)	$E_1 = 0.0001(WBC)^3 + 18.711(Sex)^2$, $E_2 = 1.469(Age)$	0.188	1.757
Model 5 for (MCV, sex and age)	$E_1 = 18.711(Sex)^2$, $E_2 = 1367.932(MCV)^{1/4} + 1.469(Age)$	0.215	1.699
Model 6 for (HCT, sex and age)	$E_1 = 18.711(Sex)^2$, $E_2 = -60.591(HCT) + 1.469(Age)$	0.366	1.372
Model 7 for (MCHC, sex and age)	$E_1 = 18.711(Sex)^2$, $E_2 = -297.972(MCHC)^{1/2} + 1.469(Age)$	0.216	1.697
Model 8 for (PLT, sex and age)	$E_1 = 18.711(Sex)^2$, $E_2 = -26.450(PLT)^{1/3} + 1.469(Age)$	0.196	1.739
Model 9 for (HB, RBC, sex and age)	$E_1 = 0.001(HB)^6 + 0.014(RBC)^5 + 18.711(Sex)^2$,$E_2 = 1.469(Age)$	0.555	0.964
Model 10 for (MCH, WBC, sex and age)	$E_1 = -0.001(MCH)^4 + 0.0001(WBC)^3 + 18.711(Sex)^2$, $E_2 = 1.469(Age)$	0.000	4.397
Model 11 for (MCV, HCT, sex and age)	$E_1 = 18.711(Sex)^2$, $E_2 = -60.591(HCT) + 1367.932(MCV)^{1/4} + 1.469(Age)$	0.371	1.364
Model 12 for (MCHC, PLT, sex and age)	$E_1 = 18.711(Sex)^2$, $E_2 = -297.972(MCHC)^{1/2} - 26.450(PLT)^{1/3} + 1.469(Age)$	0.261	1.602
Model 13 for (WBC, MCV, HCT, MCHC, sex and age)	$E_1 = 0.0001(WBC)^3 + 18.711(Sex)^2$, $E_2 = -60.591(HCT) - 297.972(MCHC)^{1/2} + 1367.932(MCV)^{1/4} + 1.469(Age)$	0.381	1.348
Model 14 for (HB, RBC, MCH, PLT, sex and age)	$E_1 = 0.001(HB)^6 + 0.014(RBC)^5 - 0.001(MCH)^4 + 18.711(Sex)^2$, $E_2 = -26.450(PLT)^{1/3} + 1.469(Age)$	0.671	0.715

Table 4.4 Optimization of the residual sum of squares to estimate the parameters by the regression optimization

Iteration Number		1	50	100	122	141	171
Residual Sum of Squares		2457.001	2422.585	1493.893	271.296	271.148	271.148
Constant	b_0	68.912	5.888	213375.919	320324.494	45059.962	2489.986
HB	b_1	-234.704	0.259	0.062	0.161	0.023	0.001
RBC	b_2	-211.649	-0.366	0.290	1.821	0.259	0.014
MCH	b_3	-80.854	0.759	0.762	-0.083	-0.012	-0.001
WBC	b_4	-689.697	-0.069	-0.068	-0.007	-0.001	-0.00005
MCV	b_5	-991415.201	-3405.904	-1583.487	2270.951	339.312	18.711
HCT	b_6	-80326.882	-226.747	96.364	-7720.240	-1098.810	-60.591
MCHC	b_7	-305965.063	-860.494	430.136	-39602.005	-5400.125	-297.972
PLT	b_8	-266546.846	-761.492	162.005	-3493.601	-477.549	-26.450
Sex	b_9	-531827.894	-1500.534	622.720	178493.976	24782.155	1367.932
Age	b_{10}	-82254.965	-239.766	230.978	168.767	26.561	1.469

Table 4.5 Comparison of the results of the multiple nonlinear regression with the two methods

Methods	SSE	MSE	R^2
Nonlinear Regression Analysis	271.148	0.514	0.766
Nonlinear Deep Learning Methods (LSTM)	273.465	0.560	0.760
Nonlinear Regression Neural Networks	287.826	0.534	0.752

LSTM: Long Short Term Memory

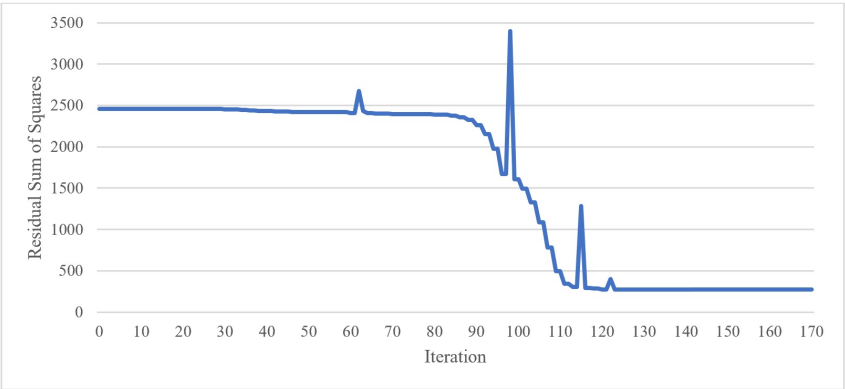


Figure 4.3 The behaviour of the residual sum of square errors by the regression optimization when the iteration is 171

4.5 Discussion and Analysis

Nonlinear models play a very important role in life since many natural phenomena such as biology, medicine, chemistry, physics, and others are represented by the corresponding models. In this respect, the necessity of a nonlinear model to predict types of anemia is revealed. Therefore, it is here concentrated on the prediction of types of anemia through the biomedical parameters under the consideration of the fitted nonlinear model.

In the literature, many researchers [92, 94, 95, 97, 103] have considered anemia problems at various levels, through the multiple nonlinear regression analysis. Note that researchers in the literature [37, 38, 89, 90, 98] used a very limited number of blood variables or a few anemia types and various nonlinear models for the investigation of different diseases and did not pay enough attention to the relationship between such a very large number of blood variables and those types of anemia. Here it is focused on a nonlinear model investigating the relationship between a large number of the blood variables and the types of anemia.

At the same time, other nonlinear models to the data have been considered, and the most accurate results for the model (4.12) have been found. These variables consist of HB, RBC, MCH, WBC, MCV, HCT, MCHC, PLT, sex, and age. The current analysis was considered to fit a nonlinear model presenting the link between observational variables (blood variables, age, and sex) and the types of anemia. Several types of models, based on the variables, are derived (see Table 4.3) in terms of less number of blood variables for the prediction of the types of anemia. The model produced in equation (4.12), see Table 4.2, in terms of a larger number of blood variables show a better correlation than the models produced in Table 4.3. Therefore, this study here concentrates on the discovery of the relationship between as large as possible, more realistic, blood variables and the types of anemia through the current model.

The model has been seen to be significantly effective on the prediction of the types of anemia ($R^2 = 0.766$). The model explains 76.60% of the change in the relationship between the observational variables and the types of anemia. That is, as realized from Table 4.2, all the variables used have a significant effect on the model.

The model uses the mean square residuals to measure the extent of distributing the

random error around the model, $MSE = 0.514$ (see Table 4.2). Small values of the MSE indicate the concentration of data around the regression line. The MSE indicates that the residuals are naturally distributed. In addition, the multiple regression uses the sum of square errors to measure how well the data fit the model. Thus, $SSE = 271.148$ means that the data fits for the model. One can notice that the current model ($R^2 = 0.766, MSE = 0.514, SSE = 271.148$). Therefore, the model is seen to be more realistic in predicting the effect of blood variables, age, and sex on the types of anemia. The values **B** refer to the estimated parameter values of the real parameters obtained by the regression optimization when the SSE indicating the estimated residual sum of squares value, which started at 2456.997 and through the optimization, it has been obtained as 271.148 (see Table 4.4 and Figure 4.3). Estimating the parameters of the model is a difficult task for classical algorithms for improvement. The starting values of the parameters have been selected. Therefore, the regression technique has been taken to obtain the optimum solution. In the estimation process, the residual sum of squares reaches its level of stability at the iteration of 141.

In the current study, it was observed to the results of the regression analysis better than Neural Networks (see Table 4.5). This is because the nonlinear regression model is easy to implement and expected to provide optimum estimates. Additionally, the regression model is a special neural network model with no hidden layers, that is, consisting of just a single neuron, it acts upon multi-inputs to produce one output. So, we can compute the optimal regression model directly and efficiently. The ANNs cannot compute an optimal model directly, when adding an activation function and possibly hidden layers. In this case, there are no guarantees that the process will converge, or that we will find the best model. It is also a lot slower than the direct solution. So, in the regression analysis, we are forced to use an iterative solution: an algorithm that goes through steps, usually improving the model with each step.

This chapter addresses the anemia forecasting issue by the nonlinear regression in comparing with two rival methods, the nonlinear deep learning method (LSTM) and the nonlinear regression neural network [105]. The computed results reveal that the multiple nonlinear regression has the best fit to the initial dataset comparing to the two competitors (see Table 4.5). Thence, this study presents a relatively very accurate nonlinear model for predicting anemia types. Additionally, since the convergence be-

havior of the nonlinear regression analysis is shown that it is in a rapid convergence tendency, reaches its level of stability at the iteration of 141 and the iteration number is limited to 171 iterations (see Table 4.4 and Figure 4.3).

4.6 Conclusions

Multiple nonlinear regression model, for the first time, has been derived in predicting the anemic diseases. The parameter values produced have all been seen to be the optimum values obtained from the multiple nonlinear regression approach, to find the approach that is more realistic. It has also been seen that the proposed multiple nonlinear regression method has a very rapid convergence tendency. The results confirm that the multiple nonlinear regression model is adequate and has a high ability to predict. In the analysis of the current anemia problem, the multiple nonlinear regression method has been found to be more accurate than nonlinear deep learning methods and nonlinear regression neural network. It has been concluded that the model is expected to be helpful for diagnosis of the types of anemia to health providers and designing appropriate treatment programs for their patients.

5

PARAMETER ESTIMATION TO ANEMIA MODELS USING THE PARTICLE SWARM OPTIMIZATION

5.1 Introduction

The progress of medical models considered to produce medical outputs is important tools to deal with the behavior of a medical problem. They depend on the quality of any particular objective achieved on the state of knowledge about the system and how well successful modeling. As indicated in the literature [10, 12, 83, 86, 88, 106], mathematical medical modelling has been realized to be a fundamentally important tool for the analysis of pathological characteristics. Response to a medical model to limits of performance is of major interest and thus the current medical model describes the relationship, between the biomedical variables and the diseases. The observational data may be modelled by a function linearly. Here the parameters for each of the variables in the linear medical model are estimated that to be the optimal model for more accurate prediction of anemia through the biomedical information.

Many models have been produced in dealing with various medical problems in the literature such as congenital heart disease [107], diabetic nephropathy [108], osteoporosis [109], and cancers [110, 111]. A frequently encountered medical problem is that of having a set of data, which one wishes to describe it by a mathematical model and determine a set of parameters that characterize the model. In this study, the major emphasis will be the fitting parameters of the model assumed to have some particular medical or mathematical significance through estimating best values in the set of the parameters. Therefore, the main aim here is to develop a medical model to study the effect of the blood variables, sex, and age on the pathologies through a large group of the variables because there has been an increase in the incidence of anemia among

different segments of society.

Some other estimation methods [35, 112, 113, 114] to analyze disease problems in addition to anemia. Heuristic algorithms can be effectively used to find the optimal parameters for the linear model in plenty of medical studies. Therefore, the PSO is one of the most efficient optimization algorithms that are used for a wide range of complex optimization problems. In computational science, the PSO is a computational method that works to improve the problem by repeatedly trying to improve the candidate solution. Therefore, these candidate solutions are created by the method repeatedly for improving the possibility of being the actual solution.

The PSO inspired by the behaviour of social models for flocking birds or fish education are based on individual improvement and social collaboration [78, 115, 116, 117, 118]. In this study, the PSO approach has been proposed to estimate the best parameter values of the linear medical model. This algorithm is common in the academic community as a typical tool because of its ability to optimize complex search spaces. Thus, the above advantages of the PSO sent us to use in dealing with the current medical problem. It should be borne in mind that fewer blood variables may cause the problem not to be effectively represented.

This chapter is structured as follows. The next section discusses the study samples of the medical dataset, explain the models procedure, the PSO algorithm, and how to test the model. Section 5.3 estimate parameters of the medical models. Section 5.4 presents the results and discussion. Finally, conclusions and recommendation for future work have been detailed.

5.2 Materials and Methods

5.2.1 Study Samples of the Medical Dataset

The data used here were collected from observations of anemia and included (539 subjects, 211 healthy subjects, 328 sick subjects) provided from blood laboratories in Iraq and we have taken observations of the ages of individuals between (6-56) years. Here, we have some blood diseases are Iron deficiency anemia (1), Deficiency Vitamin B12 (2), Thalassemia (3), Sickle cell (4) and Spherocytosis (5). For each disease, we have samples for the individuals and for each individual readings of the blood vari-

ables are Hemoglobin (HB), Red Blood Cell (RBC), Mean Corpuscular Hemoglobin (MCH), White Blood Cell (WBC), Mean Corpuscular Volume (MCV), Haematocrit (HCT), Mean Corpuscular Hemoglobin Concentration (MCHC), Platelets (PLT), and sex (male (1) and female (2)), and age. The number of variables studied for the model is consisting of ten independent variables and a dependent variable. The dependent variable consists of six different types of output (healthy subject: 0 and blood diseases: 1-5).

5.2.2 Modelling

5.2.2.1 Linear Model

A linear model is an engine behind a multitude of data applications used for many forms of prediction. Therefore, processes are governed by linear models in various fields of science such as the estimation of the parameters of a linear medical model for predicting anemia.

A linear medical model describes a linear relationship between the dependent and independent variables. The derived model is as follows:

$$\mathbf{y} = B_0 + B_1 x_1 + B_2 x_2 + ... + B_k x_k + \epsilon = B_0 + \sum_{i=1}^{k} B_i x_i + \epsilon. \tag{5.1}$$

The linear model with k predictor variables and the observations recorded for each of these n levels can be expressed in the following style

$$\begin{aligned}
y_1 &= B_0 + B_1 x_{11} + B_2 x_{12} + ... + B_k x_{1k} + \epsilon_1 \\
y_2 &= B_0 + B_1 x_{21} + B_2 x_{22} + ... + B_k x_{2k} + \epsilon_2 \\
&\vdots \\
y_i &= B_0 + B_1 x_{i1} + B_2 x_{i2} + ... + B_k x_{ik} + \epsilon_i \\
&\vdots \\
y_n &= B_0 + B_1 x_{n1} + B_2 x_{n2} + ... + B_k x_{nk} + \epsilon_n
\end{aligned} \tag{5.2}$$

Here $y_1, y_2, ..., y_n$, and $x_1, x_2, ..., x_k$, stand for the dependent and independent observations, respectively.

System (5.2) can be reexpressed in a more compact way:

$$\mathbf{y} = \mathbf{BX} + \epsilon, \tag{5.3}$$

with

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix}, \mathbf{B} = \begin{bmatrix} B_0 \\ B_1 \\ \vdots \\ B_k \end{bmatrix}, \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \tag{5.4}$$

where **y**,**X**,**B** and ϵ indicate to the observations, the parameters of the model and the unobserved random variable that adds noise to the linear relationship, respectively. To obtain the linear model, **B** should be known. **B** is estimated by minimizing the sum of the squared error function $SSE(\mathbf{B})$ under the consideration of the PSO. Knowing the estimates $\hat{\mathbf{B}}$, the linear model can now be expressed as [66, 67]

$$\hat{\mathbf{y}} = \hat{\mathbf{B}}\mathbf{X}, \tag{5.5}$$

where $\hat{\mathbf{y}}$ is the estimated value for **y**.

5.2.2.2 Nonlinear Model

Nonlinear models are important tools because life is nonlinear and many physical processes and natural phenomena encountered in the physical environment such as biology, medicine, chemistry, physics, and other areas are better represented by a nonlinear model. Therefore, most processes are governed by nonlinear models in various fields of science such as the estimation of the parameters of a nonlinear medical model for predicting the anemia types.
A nonlinear model can be given in a basic form,

$$y = f(x, b) + \epsilon, \tag{5.6}$$

where $y, x, b, f()$ and ϵ indicate the observations, the vector of the coefficients, the known nonlinear function and the unobserved random variable that adds noise to the

nonlinear relationship, respectively.

A nonlinear medical model describes the relationship between the dependent and independent variables when the behaviour of the model is nonlinear. Many nonlinear models can be proposed for dealing with anemia problem. In this regard, here, several attempts have been made to obtain the best results, depending on the biomedical variables and the exponent of each input variable of interest. Other types of nonlinear models for the data were also considered and it was noted that the current model was the best among these models to obtain accurate results.

$$y = \frac{b_0}{E_1 + E_2} + \epsilon \tag{5.7}$$

Here the denominator of equation (5.7) is separated into two parts as E_1, E_2. Thus, the separated parts are clearly expressed as:

$E_1 = b_1(HB)^6 + b_2(RBC)^5 + b_3(MCH)^4 + b_4(WBC)^3 + b_5(Sex)^2$
$E_2 = b_6(HCT) + b_7(MCHC)^{\frac{1}{2}} + b_8(PLT)^{\frac{1}{3}} + b_9(MCV)^{\frac{1}{4}} + b_{10}(Age)$.

Here y is the type of anemia, $b_i, 0 \leq i \leq 10$, are the parameters to be determined.
Here HB, RBC, MCH, WBC, MCV, HCT, MCHC, PLT indicate Hemoglobin, Red Blood Cell, Mean Corpuscular Hemoglobin, White Blood Cell, Mean Corpuscular Volume, Haematocrit, Mean Corpuscular Hemoglobin Concentration, Platelets, respectively.

5.2.3 Particle Swarm Optimization

The PSO is a population-based stochastic approach, invented by Eberhart and Kennedy [72], for solving continuous and discrete problems. They inspired from social behavior of bird flocking or fish schooling, these animals have a major role in the development of the algorithm.

The method optimizes a problem by trying to improve a solution. Each particle traces its coordinates in the area of the problem that relates to the best solutions carried out so far. This value is called *pbest*. Another "best" value that is tracked by the PSO is the best value, obtained so far by any particle in the neighbors of the particle. This location is called *lbest*. When the particle considers the whole population as its topological neighbors, the best value is a global best and is called *gbest*. The PSO idea consists of, at each time step, changing the velocity of each particle towards the *pbest* and *lbest*

locations.

In the PSO, simple software agents, called particles, move in the search space for improvement. These randomly selected particles search solution space using the information of their neighborhood, personal information, and randomness. The position of a particle represents a candidate solution to the existing improvement problem. All particles look for better sites in the search space by changing their velocity at the end of each iteration. Because of each iteration, the position and velocity vectors are expressed as follows:

$$V_i^{t+1} = \omega V_i^t + c_1 r_1 (P_{best} - X_i^t) + c_2 r_2 (G_{best} - X_i^t) \tag{5.8}$$

$$X_i^{t+1} = X_i^t + V_i^{t+1} \tag{5.9}$$

where $t, \omega, c_1, c_2, r_1, r_2, V_i^t, X_i^t, P_{best}$ and G_{best} indicate iteration number, weight parameter, acceleration coefficients (cognitive parameter, social parameter), random numbers uniformly distributed between 0 and 1, velocity of individual i at iteration t, position of individual i at iteration t, the best local value of each particle, the best value of swarm, respectively [78, 79, 80].

5.2.4 Test for the Model

The coefficient of the determination, usually referred to as R^2, is a measure explaining the change in the relationship between all blood variables, sex, and age and the anemia types.

Here, we present some initial considerations. Consider the variance of the observations y by analyzing the total sum of squares, denoted by SST and the sum of squared errors, denoted by SSE. That is,

$$SST = \sum_{j=1}^{n} (y_j - \bar{y})^2, \tag{5.10}$$

and

$$SSE = \sum_{j=1}^{n} (y_j - \hat{y}_j)^2 = \sum_{j=1}^{n} e_j^2. \tag{5.11}$$

Now, the coefficient of the determination is defined by

$$R^2 = \frac{SST - SSE}{SST}. \tag{5.12}$$

If the percentage explained by the coefficient of the determination is small, compatibility may not be very appropriate.

A terminological difference arises in the expression root mean squared error (RMSE). It is the square root of the average squared differences between the prediction and actual observations. The RMSE indicate the concentration of data around the model. In other words, it tells us how the data is centered around the most appropriate line [66, 67, 70]. It is very common to use the RMSE in the predictions. Then it is given by

$$RMSE = \sqrt{MSE}. \tag{5.13}$$

Thus, it is given by

$$MSE = \frac{1}{n}\sum_{j=1}^{n} e_j^2. \tag{5.14}$$

5.3 Estimation of the Parameters of the Model

5.3.1 Linear Model

The currently linear medical model is a linear equation for our data. The model is as follows:

$$\begin{aligned} \mathbf{y} =& B_0 + B_1 HB + B_2 RBC + B_3 MCH + B_4 WBC + B_5 MCV \\ &+ B_6 HCT + B_7 MCHC + B_8 PLT + B_9 Sex + B_{10} Age + \epsilon \end{aligned} \tag{5.15}$$

where $\mathbf{y}$ is the type of anemia and B_i, $0 \leq i \leq 10$, are the parameters to be determined. Here HB, RBC, MCH, WBC, MCV, HCT, MCHC, PLT stand for Hemoglobin, Red Blood Cell, Mean Corpuscular Hemoglobin, White Blood Cell, Mean Corpuscular Volume, Haematocrit, Mean Corpuscular Hemoglobin Concentration, Platelets, respectively.

As previously mentioned, the model can be represented in a more compact form as follows:

$$\hat{\mathbf{y}} = \hat{\mathbf{B}}\mathbf{X} \tag{5.16}$$

where

$$\hat{\mathbf{y}} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{539} \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & HB_{11} & RBC_{12} & \dots & Age_{110} \\ 1 & Hb_{21} & RBC_{22} & \dots & Age_{210} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & HB_{539,1} & RBC_{539,2} & \dots & Age_{539,10} \end{bmatrix}, \mathbf{B} = \begin{bmatrix} B_0 \\ B_1 \\ \vdots \\ B_{10} \end{bmatrix} \tag{5.17}$$

Here $\hat{\mathbf{y}}$, $\mathbf{X}$ and $\hat{\mathbf{B}}$ represent the estimates for output (anemia), the independent observations matrix, and estimated parameters, respectively.

This study aims at estimating the parameters $\mathbf{B}$ by minimizing the sum of the squared error function $SSE(\mathbf{B})$ under the consideration of the PSO.

Hence, the fitness function in the PSO search engine is selected as the $SSE(\mathbf{B})$, specifically:

$$SSE(B) = \sum_{i=1}^{n} (y_i - f(x_i, B))^2. \tag{5.18}$$

For the linear model in equation (5.15),

$$\begin{aligned} SSE(B) = \sum_{i=1}^{539} [y_i - (B_0 + B_1 HB + B_2 RBC + B_3 MCH + B_4 WBC + B_5 MCV \\ + B_6 HCT + B_7 MCHC + B_8 PLT + B_9 Sex + B_{10} Age)]^2. \end{aligned} \tag{5.19}$$

Here y_i are the dependent observations, $B_i, 0 \leq i \leq 10$, are the parameters to be determined.

In this chapter, the PSO is effectively used to estimate the parameters of the linear medical model in deriving an accurate model by finding a rapid convergence of the minimum value of the sum of the squared error in fewer iterations provides accurate estimates for parameter estimation of the linear medical model (see Tables 5.1-5.5).

The settings for the main parameters of the PSO method (ω, c_1, c_2, and the size of the swarm) determine how to optimize the search space. Usually decreases the parameter ω from around 0.9 to around 0.4 during the computation, the appropriate value for the parameter ω provides a balance between the global and local exploration capacity of the swarm and thus a better solution [73, 116, 117, 118]. If the parameter ω is much less than one, only a small momentum of the previous time step is preserved,

thus rapid changes in the direction are possible with this setting. High settings near 1 facilitate global searching. The usual choices for acceleration coefficients are c_1 and c_2, usually, c_1 is equal to c_2 and ranges between 0 and 4. The size of swarm plays a very important role in the PSO, as is the durability and complexity of the algorithm. By inspiring from the literature [73, 117, 118], we have produced our PSO algorithm as given in Figure 5.1.

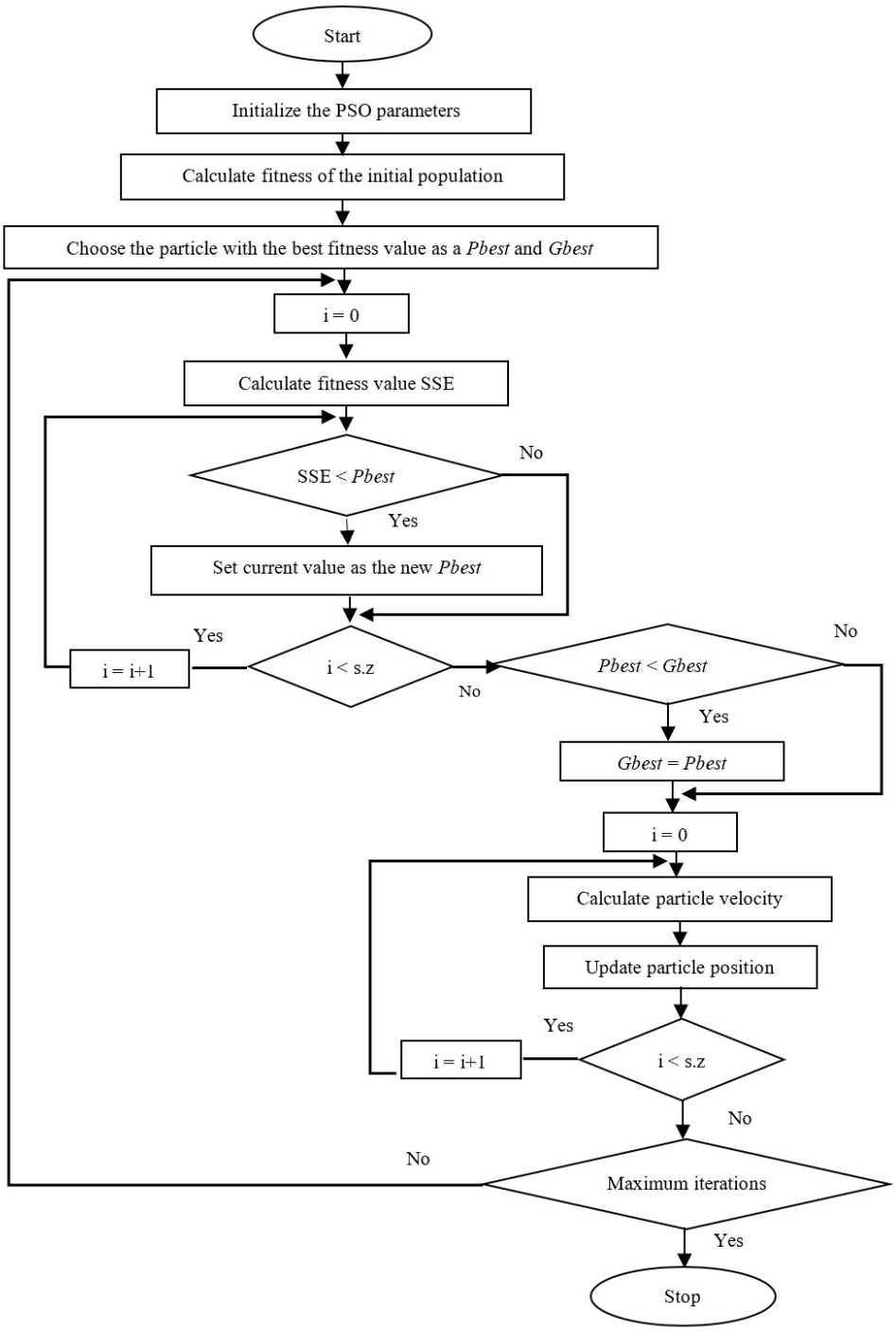


Figure 5.1 The PSO algorithm for the estimation of the parameters of the linear model

5.3.2 Nonlinear Model

The nonlinear function f in (5.6) has parameters given by $b = (b_1, b_2, \ldots)$. This study aims at estimating the parameters b obtained by minimizing the sum of the squared error function $SSE(b)$ under the consideration of the PSO.

Hence, the cost (fitness) function in the PSO search engine is selected as the $SSE(b)$, specifically:

$$SSE(b) = \sum_{i=1}^{n} (y_i - f(x_i, b))^2. \tag{5.20}$$

For instance, for the model in equation (5.7),

$$SSE(b) = \sum_{i=1}^{539} [y_i - \frac{b_0}{E_1 + E_2}]^2, \tag{5.21}$$

where $E_1 = b_1(HB)^6 + b_2(RBC)^5 + b_3(MCH)^4 + b_4(WBC)^3 + b_5(Sex)^2$
$E_2 = b_6(HCT) + b_7(MCHC)^{\frac{1}{2}} + b_8(PLT)^{\frac{1}{3}} + b_9(MCV)^{\frac{1}{4}} + b_{10}(Age)$.

Here y_i are the dependent observations, $b_i, 0 \leq i \leq 10$, are the parameters to be determined.

The main parameters of the PSO method are ω, c_1, c_2, and the size of the swarm. The settings for these parameters are decided according to how to optimize the search space. The inertia weight is used to control the effect of the previous history of velocities on the current velocity. Thus, the parameter ω regulates the trade-off between the global and local exploration capabilities of the swarm and also provides a balance between the global and local exploration capacity of the swarm and to find a better solution [73, 74, 119]. The usual choices for acceleration coefficients are cognitive parameter c_1 and social parameter c_2, usually, c_1 is equal to c_2 ranged between 0 and 4. The swarm size plays a very important role in the PSO, as is the complexity and sturdiness of the algorithm. From the literature [73, 120], we have inspired our PSO algorithm as shown in Figure 5.2.

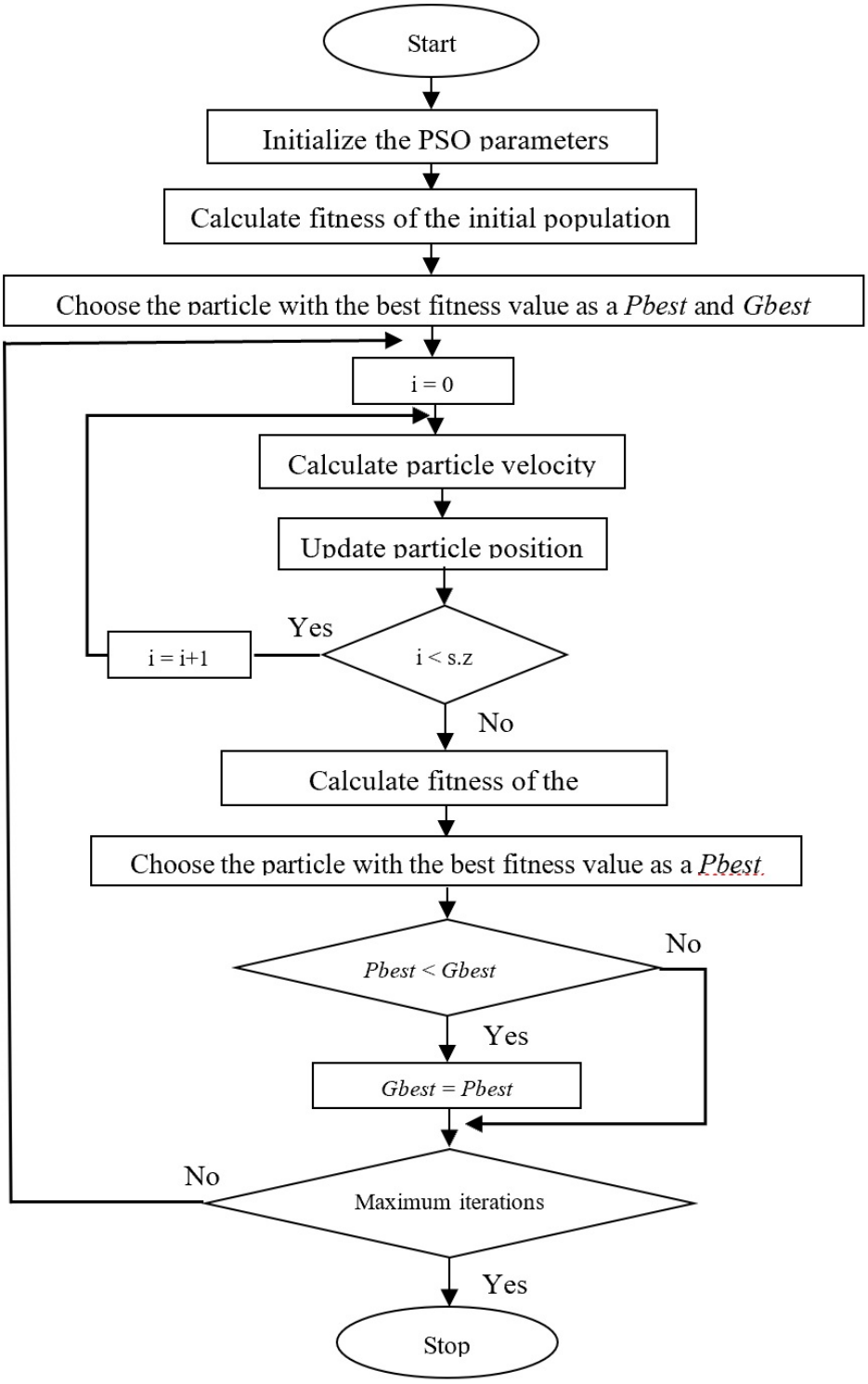


Figure 5.2 The PSO algorithm for the estimation of the parameters of the nonlinear model

5.4 Discussion

5.4.1 Linear Model

The current study focuses to obtain the best estimate of the parameters through the PSO for the currently derived linear model to detect the link between the biomedical variables and anemia.

As opposed to the PSO approach, classical methods in dealing with linear models have some disadvantages as seen in the previous works [11, 74, 75, 121, 122], where they require many mathematical operations; like the Jacobean matrix, and matrix operations.

The researchers estimated parameters of a great number of models by using the PSO in the literature [73, 76, 77, 123, 124, 125, 126]. They discussed different problems/models by using their own approaches. We have here studied a linear model for a great number of biomedical data of anemia through the PSO to estimate the parameters for the model and investigating the relationship between many blood variables and the anemia types as opposed to researchers in the literature [37, 38, 89, 90], they used a very limited number of blood variables or a few the anemia types.

Here, we have estimated the parameters of the linear model through the PSO algorithm (see Tables 5.1-5.4), and the produced results for various versions of the model by the minimum error (see Table 5.5). In the estimation, when the number of iterations is increasing, the error is decreasing as seen in Figures 5.3-5.6. Notice that the iteration reaches its optimum level at 4500.

Table 5.1 Parameter estimation by the PSO algorithm when the iteration is 500.

Biomedical Variables	Parameters $B_i, 0 \leq i \leq 10$	SST	$SSE(B)$	$RMSE$	R^2
Constant	-3.167	1157.243	1817.378	1.836	0
HB	-0.726				
RBC	-0.634				
MCH	0.901				
WBC	0.009				
MCV	0.125				
HCT	0.062				
MCHC	-1.408				
PLT	0.257				
Sex	0.465				
Age	0.010				

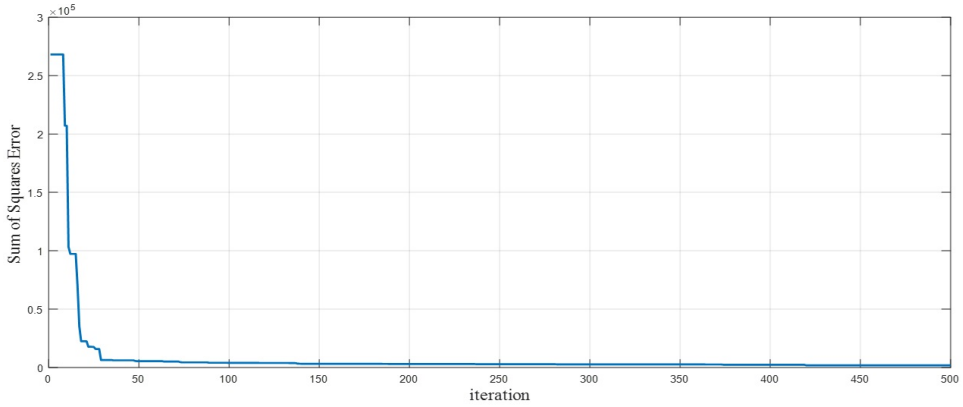


Figure 5.3 Sum of square errors of the PSO algorithm when the iteration is 500

Table 5.2 Parameter estimation by the PSO algorithm when the iteration is 1000.

Biomedical Variables	Parameters $B_i, 0 \leq i \leq 10$	SST	$SSE(B)$	$RMSE$	R^2
Constant	7.799	1157.243	1080.449	1.416	0.066
HB	1.399				
RBC	1.141				
MCH	-0.275				
WBC	-0.018				
MCV	0.114				
HCT	-0.743				
MCHC	-0.149				
PLT	0.002				
Sex	0.596				
Age	0.023				

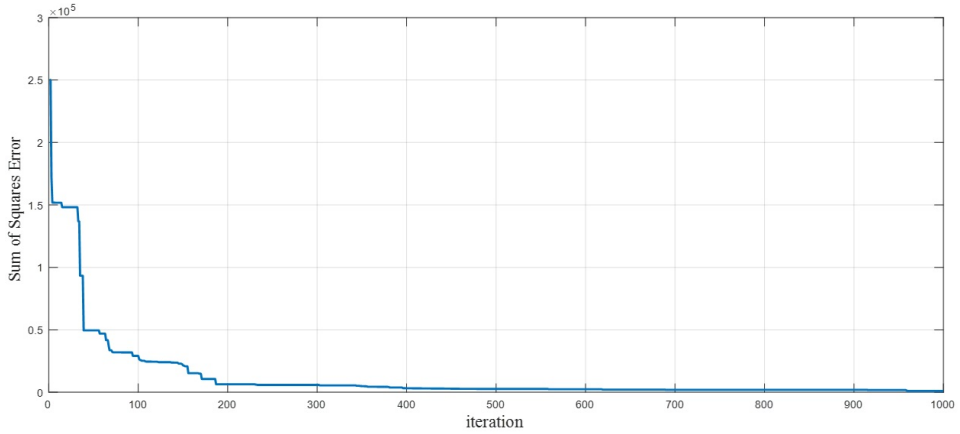


Figure 5.4 Sum of square errors of the PSO algorithm when the iteration is 1000

Table 5.3 Parameter estimation by the PSO algorithm when the iteration is 2000.

Biomedical Variables	Parameters $B_i, 0 \leq i \leq 10$	SST	$SSE(B)$	$RMSE$	R^2
Constant	5.603	1157.243	405.983	0.868	0.649
HB	0.252				
RBC	0.270				
MCH	0.146				
WBC	0.0002				
MCV	0.053				
HCT	-0.304				
MCHC	-0.038				
PLT	0.0008				
Sex	-0.208				
Age	-0.013				

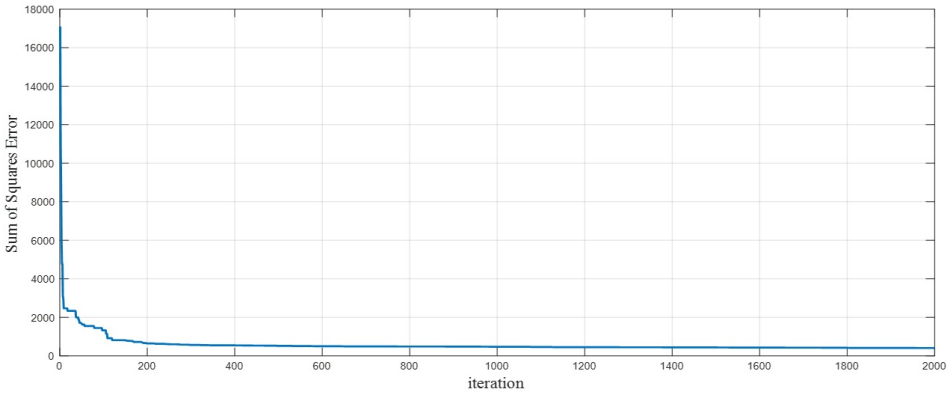


Figure 5.5 Sum of square errors of the PSO algorithm when the iteration is 2000

Table 5.4 Parameter estimation by the PSO algorithm when the iteration is 4500.

Biomedical Variables	Parameters $B_i, 0 \leq i \leq 10$	SST	$SSE(B)$	$RMSE$	R^2
Constant	6.345	1157.243	347.989	0.803	0.699
HB	-0.201				
RBC	-0.461				
MCH	-0.033				
WBC	0.001				
MCV	0.003				
HCT	-0.022				
MCHC	0.003				
PLT	0.001				
Sex	-0.306				
Age	-0.009				

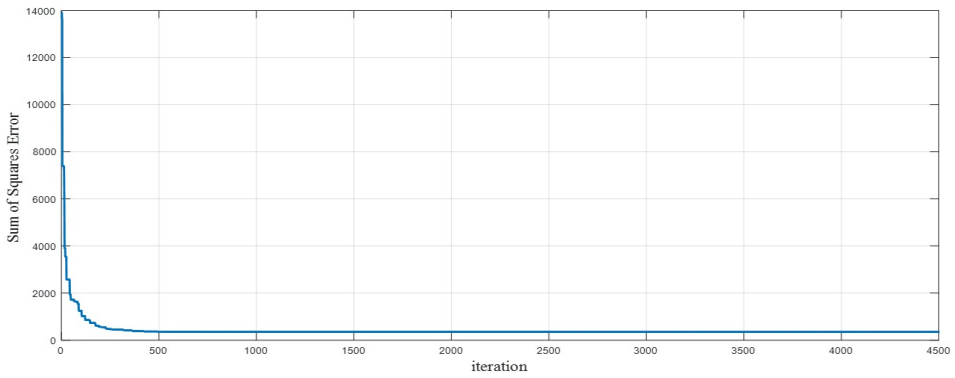


Figure 5.6 Sum of square errors of the PSO algorithm when the iteration is 4500

Table 5.5 Parameter estimation of the various forms by the PSO algorithm when the iteration is 4500.

Models	SSE	$RMSE$	R^2
Model 1 for (HB, sex and age)	500.117	0.963	0.568
Model 2 for (RBC, sex and age)	956.017	1.332	0.174
Model 3 for (MCH, sex and age)	862.084	1.265	0.255
Model 4 for (WBC, sex and age)	891.756	1.286	0.229
Model 5 for (MCV, sex and age)	937.336	1.319	0.190
Model 6 for (HCT, sex and age)	406.077	0.868	0.648
Model 7 for (MCHC, sex and age)	876.008	1.275	0.243
Model 8 for (PLT, sex and age)	843.894	1.251	0.271
Model 9 for (HB, MCH, sex and age)	496.170	0.959	0.571
Model 10 for (RBC, WBC, sex and age)	885.520	1.282	0.235
Model 11 for (MCV, PLT, sex and age)	829.614	1.241	0.283
Model 12 for (MCHC, HCT, sex and age)	389.654	0.850	0.663
Model 13 for (HB, WBC, HCT, sex and age)	384.303	0.844	0.667
Model 14 for (MCV, MCHC, RBC, sex and age)	844.280	1.252	0.270
Model 15 for (HB, RBC, MCH, WBC, sex and age)	353.664	0.810	0.690
Model 16 for (MCV, HCT, MCHC, PLT, sex and age)	378.580	0.838	0.670

In this study, the size of the swarm is taken to be according to the structure of the linear medical model, the number of estimated parameters, and searching space between (-10 and 10). The acceleration coefficients; cognitive parameter c_1 and social parameter c_2 are selected as 1 and 3, respectively. The algorithm is set to stop after different iterations and different independent experiments to check the durability of the estimation strategy.

Estimating the parameters of the medical model is a difficult task for classical methods of optimization. The starting values for the parameters are randomly selected from the search area. The **B** values refer to the estimated parameter values for the real parameters obtained by the PSO. After different independent attempts have been made and different iterations 500, 1000, 2000 and 4500 have been taken to obtain the best parameters, and then we have obtained the best estimated parameters with iterations of 4500 (see Tables 5.1-5.4 and Figures 5.3-5.6).

Since the PSO algorithm is random inherently, convergence behavior and final estimated values can be of attention. For the medical model, the behavior of the error

function is interpreted through the PSO approach, which consists of the values evaluated during the process of minimization (see Figures 5.3-5.6).

The parameter value is suitable for the model, when $SSE = 347.989, RMSE = 0.803$,and $R^2 = 0.699$ by the PSO. This is important because the SSE measures how well the data fit the model and means a better fit the model with the data and small values of the RMSE indicate the concentration of data around the model line. The medical model of interest has been seen to be effective significantly, on the prediction of the anemia types, which explain 69.90% of the change in the relation of the model between the observational variables and the anemia types.

The results obtained from the SSE,RMSE,and R^2 by using the PSO at the iteration of 4500, that the models produced in terms of a great number of blood variables a better relationship appear than the models produced in terms of fewer number of blood variables for predicting the anemia types (see Tables 5.4,5.5).

5.4.2 Nonlinear Model

The current study concentrates on getting the best estimation of the parameters through the PSO for the currently derived nonlinear medical model to discover the effect of the blood variables, sex, and age on the anemia types. Thus, the parameters of the nonlinear medical model are estimated through the PSO algorithm (see Table 5.6), and the produced results for various versions of the models through the minimum error are illustrated in (Table 5.7). In the estimation step, when the number of iteration increases, the error is decreasing as seen in Figures 5.7-5.9. Notice that the iteration reaches its optimum level at 3000. It is important to note that, nonlinear regression analysis, the nonlinear deep learning (LSTM) and nonlinear regression neural network methods have also been applied to compare our model results. The results detected that the currently derived model is better than the other competitors (see Table 5.8).

Table 5.6 Estimation of the parameters of the nonlinear medical model by the PSO algorithm

Iteration Number		100	500	1000	3000
MSE		2.071	0.717	0.510	0.503
SST		1157.243	1157.243	1157.243	1157.243
SSE(b)		1116.277	386.518	275.358	271.148
R^2		0.036	0.666	0.762	0.766
Constant	b_0	-283.184	-211.634	-439.808	-337.966
HB	b_1	-0.001	-0.0001	-0.0001	-0.0002
RBC	b_2	-0.025	-0.0001	0.029	-0.002
MCH	b_3	0.0001	0.0001	0.0003	-0.00009
WBC	b_4	-1.177	0.0001	0.000006	0.00002
MCV	b_5	-9.257	-1.456	1.006	-2.600
HCT	b_6	0.0001	-0.895	-7.688	-8.733
MCHC	b_7	-10.285	1.720	-72.156	-40.975
PLT	b_8	-2.359	-0.673	-9.352	-3.851
Sex	b_9	-4.206	1.551	-115.177	-191.734
Age	b_{10}	-2.024	-0.909	0.159	-0.256

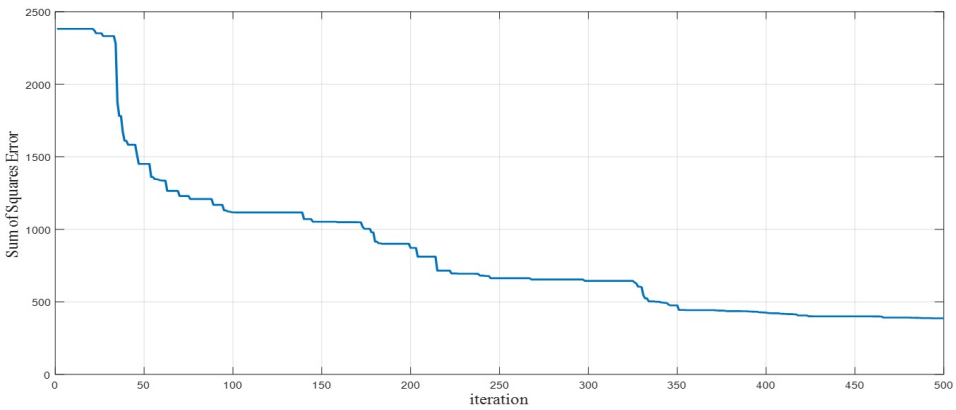


Figure 5.7 Behaviour of the sum of square errors by the PSO when the iteration is 500

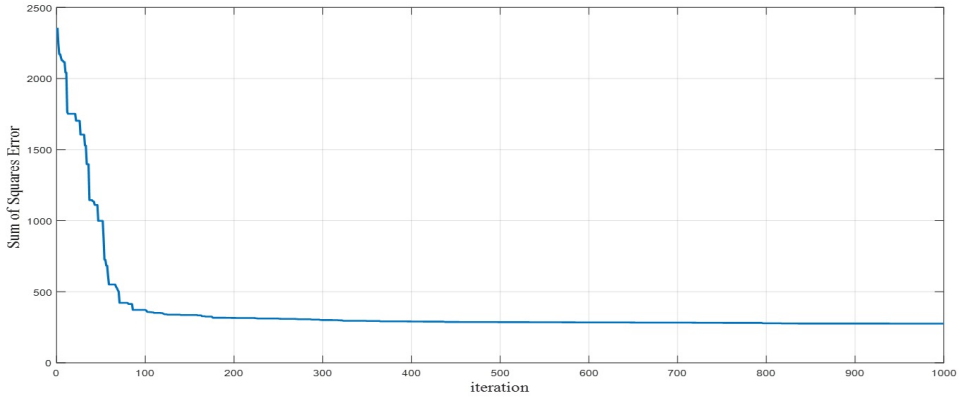


Figure 5.8 Behaviour of the sum of square errors by the PSO when the iteration is 1000

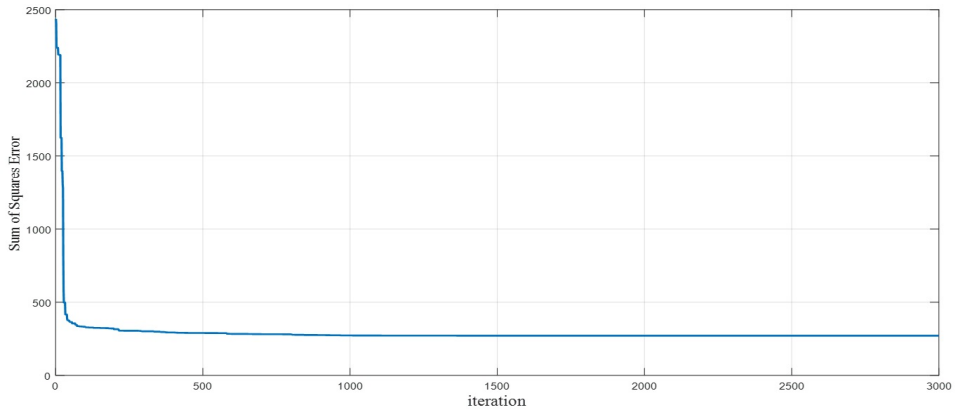


Figure 5.9 Behaviour of the sum of square errors by the PSO when the iteration is 3000

Table 5.7 Parameters Estimation of the nonlinear medical model by the PSO algorithm in various forms

Models	SSE	R^2	MSE
Model 1 for (HB, sex and age)	519.588	0.551	0.971
Model 2 for (RBC, sex and age)	940.242	0.188	1.744
Model 3 for (MCH, sex and age)	889.922	0.231	1.650
Model 4 for (WBC, sex and age)	939.826	0.188	1.757
Model 5 for (MCV, sex and age)	908.792	0.215	1.699
Model 6 for (HCT, sex and age)	734.005	0.366	1.372
Model 7 for (MCHC, sex and age)	907.781	0.216	1.697
Model 8 for (PLT, sex and age)	930.375	0.196	1.739
Model 9 for (HB, RBC, sex and age)	514.533	0.555	0.964
Model 10 for (MCH, WBC, sex and age)	882.851	0.237	1.638
Model 11 for (MCV, HCT, sex and age)	728.407	0.371	1.364
Model 12 for (MCHC, PLT, sex and age)	855.350	0.261	1.602
Model 13 for (WBC, MCV, HCT, MCHC, sex and age)	716.885	0.381	1.348
Model 14 for (HB, RBC, MCH, PLT, sex and age)	380.217	0.671	0.715

Table 5.8 Comparison of the PSO results with the other methods

Methods	SSE	MSE	R^2
PSO	271.148	0.503	0.766
Nonlinear regression analysis	271.148	0.514	0.766
Nonlinear Deep Learning Methods (LSTM)	273.465	0.560	0.760
Nonlinear Regression Neural Networks	287.826	0.534	0.752

LSTM: Long Short Term Memory

Researchers of previous studied have used a very limited number of blood variables or a few types of anemia [37, 38, 90] to investigate various diseases, and they kept the number of observational variables and the anemia types in their studies very modest as opposed to the current study. Therefore, here, we focus on an optimum nonlinear medical model investigating the relationship between many blood variables and types of anemia.

As opposed to the PSO approach, classical ways in dealing with the nonlinear model have some disadvantages as seen in the previous works [11, 74, 75, 121, 122] with

required a lot of cumbersome operations like matrix operations, gradient operations, and the Jacobean matrix. In the past [73, 74, 76, 77, 126, 127, 128], researchers estimated the parameters of a large number of various models by using the PSO. In the corresponding literature, they discussed various models/problems by using their own approaches. We have here studied a nonlinear model for a large amount of medical data on anemia to estimate the parameters of the model through the PSO.

In this work, the swarm size is calculated according to the structure of the nonlinear medical model, the number of estimated parameters, and the search space (-1000 and 10). The algorithm parameters c_1 and c_2 were selected as 1 and 3, respectively. The termination criterion was defined as the iteration limit. Specifically, the algorithm was set to stop after different independent experiments for different iterations to verify the robustness of the estimation strategy.

Estimating the parameters of the nonlinear medical model for improvement is a complicated task for classical algorithms. The b values refer to the estimated parameter values for the real parameters that the PSO obtains after randomly specifying the initial parameters of the model from the search space. After making different independent attempts and different iterations, 100, 500, 1000, and 3000 were taken to get the best parameters, and we achieved that goal at 3000 iterations (see Table 5.6 and Figures 5.7-5.9).

Since the PSO algorithm is inherently random, the behavior of convergence and the final estimated parameters values can be of interest. For the nonlinear medical model, the behavior of the error function is explained by the PSO approach which consists of the values estimated during the minimization process (see Figures 5.7-5.9). If Figures 5.7-5.9 are examined closely, the superiority variation to estimation accuracy for the parameter values of the medical model when $SSE = 271.149, MSE = 0.503$,and $R^2 = 0.766$ by the PSO may be seen. This is important because the SSE and the MSE measure how well the data fit the model and anemia types, and concentration of data around the model line, which means a better fit for the model with the data. The model has been seen to be significantly effective on the prediction of anemia types, and the model explains 76.60% of the change in the relationship between the observational variables and the anemia types.

From the results obtained in Tables 5.6 and 5.7, we see through the SSE,MSE,and

R^2 by using the PSO that the models produced in terms of a larger number of blood variables show a better correlation than the models produced in terms of fewer blood variables for predicting anemia types at the iteration of 3000.

This study addresses the anemia prediction issue by the PSO compared to other methods including nonlinear regression analysis, the nonlinear deep learning method (LSTM), and the nonlinear regression neural network. The computed results showed that the PSO has the best fit to the initial dataset compared to the others (see Table 5.8).

5.5 Conclusions

This study has discovered the anemia types through biomedical information under the consideration of eight different blood variables, sex, and age of individuals. Therefore, it has developed an alternative for estimating the parameter approach that depends on the PSO algorithm in medical models. As opposed to classical methods, it has been seen that the PSO approach is more advantageous, it requires less mathematical operations to estimate medical model parameters. It can be concluded that the PSO algorithm has been considered as an effective and very appropriate estimating method for the current and similar to current medical models. The parameter values produced are seen to be the most up-to-date and maybe the best. Thus, the PSO algorithm shows the tendency of rapid convergence for the model with the knowledge that the number of parameters is eleven.

6

RESULTS AND DISCUSSION

This chapter presents the total conclusion to the work reported in this thesis. This study has forecasted the anemia through biomedical data under the consideration of the blood variables, sex, and age of individuals. The observational blood variables are HB, RBC, MCH, WBC, MCV, HCT, MCHC, and PLT.

First, the MLR model has been derived for representing the anemia types. The results revealed that the regression model is very fitted one and is capable of representing the problem. In the analysis of the current anemia problem, the multiple regression method has been found to be slightly more accurate than linear deep learning methods.

Secondly, a multiple nonlinear regression model has been derived for representing the anemia. The parameter values produced have all been seen to be the optimum values obtained from the multiple nonlinear regression approach. It has also been seen that the proposed multiple nonlinear regression method has a very rapid convergence tendency. The results confirmed that the multiple nonlinear regression model is adequate and has a high ability to predict. The multiple nonlinear regression method has been found to be slightly more accurate than the nonlinear deep learning methods and the nonlinear regression neural network.

Thirdly, an alternative approach has been developed for estimating the parameter that relies on the PSO algorithm in the medical models. The PSO method has been used to estimate model parameters, the PSO approach presented here does not require any additional calculations. As opposed to the PSO approach, classical methods have some disadvantages because they require many intricate mathematical operations. It can be concluded that the PSO algorithm has been seen to be an effective and very suitable parameter estimation method for the current medical models. The parameter values

produced are the latest and best results for securing a more realistic approach. Thus, the PSO algorithm showed the tendency of rapid convergence for the models with the knowledge that the number of parameters is eleven. The PSO approach has been found to be more accurate than the nonlinear deep learning method and the nonlinear regression neural network.

It has been concluded that the models are expected to be helpful for the diagnosis of the anemia types to health providers and designing an appropriate treatment program for their patients. It can be accepted that the use of relatively less number of data with the current approach could have weakened importantly our results and observations. For further research, these mathematical models may be attempted to improve under the consideration of various computational methods.

References

[1] C. Judd, G. McClelland, and C. Ryan, *Data analysis: Aa model comparison approach*. Routledge, 2011.

[2] G. Wiederhold, *Database design*. McGraw-Hill, New York, 1983.

[3] W. Frawley, G. Piatetsky-Shapiro, and C. Matheus, "Knowledge discovery in databases: An overview," *AI Magazine*, vol. 13, no. 3, p. 57, 1992.

[4] A. Belle, R. Thiagarajan, S. Soroushmehr, F. Navidi, D. Beard, and K. Najarian, "Big data analytics in healthcare," *BioMed Research International*, vol. 2015, 2015.

[5] K. Priyanka and N. Kulennavar, "A survey on big data analytics in health care," *International Journal of Computer Science and Information Technologies*, vol. 5, no. 4, pp. 5865–8, 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," *Applied Artificial Intelligence*, vol. 30, no. 6, pp. 590–609, 2016.

[7] X. Liu and H. Fu, "Pso-based support vector machine with cuckoo search technique for clinical disease diagnoses," *The Scientific World Journal*, vol. 2014, 2014.

[8] L. Daniel and M. Glenn, "An introduction to mathematical modelling," *Bioinformatics and Statistics Scotland*, 2008.

[9] X. Yan and X. Su, "Linear regression analysis: theory and computing," *World Scientific*, 2009.

[10] M. Sari, E. Gulbandilar, and A. Cimbiz, "Prediction of low back pain with two expert systems. journal of medical systems," *Journal of Medical Systems*, vol. 36, no. 3, pp. 1523–7, 2012.

[11] M. Sari, C. Tuna, and S. Akogul, "Prediction of tibial rotation pathologies using particle swarm optimization and k–means algorithms," *Journal of Clinical Medicine*, vol. 7, no. 4, p. 65, 2018.

[12] M. Sari and B. Cetiner, "Predicting effect of physical factors on tibial motion using artificial neural networks," *Expert Systems with Applications*, vol. 36, no. 6, pp. 9743–6, 2009.

[13] H. Li, M. Luo, J. Zheng, J. Luo, R. Zeng, N. Feng, Q. Du, and J. Fang, "An artificial neural network prediction model of congenital heart disease based on risk factors: A hospital-based case-control study," *Medicine*, vol. 96, no. 6, 2017.

[14] N. Amma, "Cardiovascular disease prediction system using genetic algorithm and neural network," *In2012 International Conference on Computing, Communication and Applications, IEEE*, pp. 1–5, 2012.

[15] D. Sisodia and D. Sisodia, "Prediction of diabetes using classification algorithms," *Procedia Computer Science*, vol. 132, pp. 1578–85, 2018.

[16] Q. Zou, K. Qu, Y. Ju, H. Tang, Y. Luo, and D. Yin, "Predicting diabetes mellitus with machine learning techniques," *Frontiers in Genetics*, vol. 9, p. 515, 2018.

[17] J. Martínez-Martínez, P. Escandell-Montero, C. Barbieri, E. Soria-Olivas, F. Mari, M. Martínez-Sober, C. Amato, A. López, M. Bassi, R. Magdalena-Benedito, and A. Stopper, "Prediction of the hemoglobin level in hemodialysis patients using machine learning techniques," *Computer Methods and Programs in Biomedicine*, vol. 117, no. 2, pp. 208–17, 2014.

[18] M. Reymann, E. Dorschky, B. Groh, C. Martindale, P. Blank, and B. Eskofier, "Blood glucose level prediction based on support vector regression using mobile platforms," *In2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2990–2993, 2016.

[19] P. Altrock, C. Brendel, R. Renella, S. Orkin, D. Williams, and F. Michor, "Mathematical modeling of erythrocyte chimerism informs genetic intervention strategies for sickle cell disease," *American Journal of Hematology*, vol. 91, no. 9, pp. 931–7, 2016.

[20] M. Abdullah and S. Al-Asmari, "Anemia types prediction based on data mining classification algorithms," *Communication, Management and Information Technology–Sampaio de Alencar (Ed.)*, 2017.

[21] C. Yu, M. Bhatnagar, R. Hogen, D. Mao, A. Farzindar, and K. Dhanireddy, "Anemic status prediction using multilayer perceptron neural network model," *InGCAI*, pp. 213–220, 2017.

[22] M. Hasani and A. Hanani, "Automated diagnosis of iron deficiency anemia and thalassemia by data mining techniques," *International Journal of Computer Science and Network Security (IJCSNS)*, vol. 17, no. 4, p. 326, 2017.

[23] A. El-Halees and A. Shurrab, "Blood tumor prediction using data mining techniques," *Blood tumor prediction using data mining techniques*, vol. 6, 2017.

[24] T. Hamdi, J. Ali, V. D. Costanzo, F. Fnaiech, E. Moreau, and J. Ginoux, "Accurate prediction of continuous blood glucose based on support vector regression and differential evolution algorithm," *Biocybernetics and Biomedical Engineering*, vol. 38, no. 2, pp. 362–72, 2018.

[25] M. Tetschke, P. Lilienthal, T. Pottgiesser, T. Fischer, E. Schalk, and S. Sager, "Mathematical modeling of rbc count dynamics after blood loss," *Processes*, vol. 6, no. 9, p. 157, 2018.

[26] N. Meng, P. Zhang, J. Li, J. He, and J. Zhu, "Prediction of coronary heart disease using routine blood tests," *arXiv preprint arXiv*, p. 1809.09553, 2018.

[27] M. Jaiswal and A. S. T. Siddiqui, "Machine learning algorithms for anemia disease prediction," *InRecent Trends in Communication, Computing, and Electronics, Springer*, pp. 463–469, 2019.

[28] M. Lewin, S. Sarasua, and P. Jones, "A multivariate linear regression model for predicting children's blood lead levels based on soil lead levels: A study at four superfund sites," *Environmental Research*, vol. 81, no. 1, pp. 52–61, 1999.

[29] D. Makh, C. Harman, and A. Baschat, "Is doppler prediction of anemia effective in the growth-restricted fetus?," *Ultrasound in Obstetrics and Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology*, vol. 22, no. 5, pp. 489–92, 2003.

[30] M. Foster, R. Nolan, and M. Lam, "Prediction of anemia on unenhanced computed tomography of the thorax," *Journal-Canadian Association Of Radiologists*, vol. 54, no. 1, pp. 26–30, 2003.

[31] M. Vincent, G. Dranitsaris, S. Verma, C. Lau, P. Gascon, S. V. Belle, and H. Ludwig, "The development and validation of a prediction tool for chemotherapy-induced anemia in patients with advanced nonsmall cell lung cancer receiving palliative chemotherapy," *Supportive Care in Cancer*, vol. 15, no. 3, pp. 265–72, 2007.

[32] J. Schneider, M. Fujii, C. Lamp, B. Lönnerdal, K. Dewey, and S. Zidenberg-Cherr, "The use of multiple logistic regression to identify risk factors associated with anemia and iron deficiency in a convenience sample of 12–36-mo-old children from low-income families," *The American Journal of Clinical Nutrition*, vol. 87, no. 3, pp. 614–20, 2008.

[33] S. Lee, S. Cha, S. Lee, and D. Shin, "Evaluation of the effect of hemoglobin or hematocrit level on dural sinus density using unenhanced computed tomography," *Yonsei Medical Journal*, vol. 54, no. 1, pp. 28–33, 2013.

[34] J. Milton, V. Gordeuk, J. Taylor, M. Gladwin, M. Steinberg, and P. Sebastiani, "Prediction of fetal hemoglobin in sickle cell anemia using an ensemble of genetic risk prediction models," *Circulation: Cardiovascular Genetics*, vol. 7, no. 2, pp. 110–5, 2014.

[35] S. Dey and E. Raheem, "A multilevel multinomial logistic regression model for identifying risk factors of anemia in children aged 6-59 months in northeastern states of india," *arXiv preprint arXiv*, p. 1504.02835, 2015.

[36] Y. Hsieh, C. Wu, S. Lu, and Y. Tsao, "A linear regression model with dynamic pulse transit time features for noninvasive blood pressure prediction," *In2016 IEEE Biomedical Circuits and Systems Conference (BioCAS), IEEE*, pp. 604–607, 2016.

[37] Y. Chen and S. Miaou, "A kalman filtering and nonlinear penalty regression approach for noninvasive anemia detection with palpebral conjunctiva images," *Journal of Healthcare Engineering*, vol. 2017, 2017.

[38] F. Habyarimana, T. Zewotir, and S. Ramroop, "Structured additive quantile regression for assessing the determinants of childhood anemia in rwanda," *International Journal of Environmental Research and Public Health*, vol. 14, no. 6, p. 652, 2017.

[39] A. Aishah, M. Zainuriah, and A. Norhilda, "Multiple linear regression model analysis in predicting fasting blood glucose level in healthy subjects," *InIOP Conference Series: Materials Science and Engineering, IOP Publishing*, vol. 469, no. 1, p. 012050, 2019.

[40] X. Li and C. Li, "Improved ceemdan and pso-svr modeling for near-infrared noninvasive glucose detection," *Computational and Mathematical Methods in Medicine*, vol. 2016, 2016.

[41] N. Sharma, V. Khullar, and A. Luhach, "An intelligent system based on back-propagation neural network and particle swarm optimization based neural network for diagnosing anemia in pregnant ladies," *i-manager's Journal on Information Technology*, vol. 6, no. 2, pp. 27–32, 2017.

[42] J. Dai, Z. Ji, Y. Du, and S. Chen, "In vivo noninvasive blood glucose detection using near-infrared spectrum based on the pso-2ann model," *Technology and Health Care*, vol. 26, no. S1, pp. 229–39, 2018.

[43] W. H. Organization, "Worldwide prevalence of anaemia 1993-2005: Who global database on anaemia," 2008.

[44] P. Hébert, G. Wells, M. Blajchman, J. Marshall, C. Martin, G. Pagliarello, M. Tweeddale, I. Schweitzer, and E. Yetisir, "A multicenter, randomized, controlled clinical trial of transfusion requirements in critical care," *New England Journal of Medicine*, vol. 340, no. 6, pp. 409–17, 1999.

[45] A. Kim, S. Rivera, D. Shprung, D. Limbrick, V. Gabayan, E. Nemeth, and T. Ganz, "Mouse models of anemia of cancer," *PLoS One*, p. e93283, 2014.

[46] X. Li, M. Dao, G. Lykotrafitis, and G. Karniadakis, "Biomechanics and biorheology of red blood cells in sickle cell anemia," *Journal of Biomechanics*, vol. 50, pp. 34–41, 2017.

[47] N. Sirachainan, P. Iamsirirak, P. Charoenkwan, P. Kadegasem, P. Wongwerawattanakoon, W. Sasanakul, N. Chansatitporn, and A. Chuansumrit, "New mathematical formula for differentiating thalassemia trait and iron deficiency anemia in thalassemia prevalent area: a study in healthy school-age children," *Southeast Asian Journal of Tropical Medicine and Public Health*, vol. 45, no. 1, pp. 174–182, 2014.

[48] I. Roth, B. Lachover, G. Koren, C. Levin, L. Zalman, and A. Koren, "Detection of β-thalassemia carriers by red cell parameters obtained from automatic counters using mathematical formulas," *Mediterranean Journal of Hematology and Infectious Diseases*, vol. 10, no. 1, 2018.

[49] C. Jiménez, "Multiple linear regression model analysis in predicting fasting blood glucose level in healthy subjects," *Clinical Chemistry*, vol. 39, pp. 2271–2275, 1993.

[50] N. Soleimani, "Relationship between anaemia, caused from the iron deficiency, and academic achievement among third grade high school female students," *Procedia-Social and Behavioral Sciences*, vol. 29, pp. 1877–1884, 2011.

[51] S. Piplani, M. Madaan, R. Mannan, M. Manjari, T. Singh, and M. Lalit, "Evaluation of various discrimination indices in differentiating iron deficiency anemia and beta thalassemia trait: a practical low cost solution," *Annals of Pathology and Laboratory Medicine*, vol. 3, no. 6, pp. A551–559, 2016.

[52] D. Freedman, *Statistical models: theory and practice*. Cambridge University Press, 2009.

[53] R. Cook and S. Weisberg, *Criticism and influence analysis in regression*, vol. 13. Sociological Methodology, Wiley, 1982.

[54] J. Angrist and J. Pischke, *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press, 2008.

[55] F. Galton, *Presidential address, section H, anthropology*, vol. 55. British Association Reports, 1885.

[56] G. Yule, *On the theory of correlation*, vol. 60. Journal of the Royal Statistical Society, 1897.

[57] R. Fisher, "The goodness of fit of regression formulae, and the distribution of regression coefficients," *Journal of the Royal Statistical Society*, vol. 85, no. 4, pp. 597–612, 1922.

[58] S. Piplani, M. Madaan, R. Mannan, M. Manjari, T. Singh, and M. Lalit, "Fisher and regression," *Statistical Science*, vol. 20, no. 4, pp. 401–17, 2005.

[59] A. Zhang, Z. Lipton, M. Li, and A. Smola, *Dive into Deep Learning*. 2019.

[60] S. Paras, "A simple weather forecasting model using mathematical regression," *Indian Research Journal of Extension Education*, vol. 12, no. 2, pp. 161–8, 2016.

[61] K. Ho, P. Joshua, and S. Kok, "A multiple regression analysis approach for mathematical model development in dynamic manufacturing system: a case study," *Journal of Scientific Research and Development*, vol. 2, pp. 81–87, 2015.

[62] S. Amiri, M. Mottahedi, and S. Asadi, "Using multiple regression analysis to develop energy consumption indicators for commercial buildings in the us," *Energy and Buildings*, vol. 109, pp. 209–16, 2015.

[63] F. Al-Hadeethi and M. Al-Safadi, "Using the multiple regression analysis with respect to anova and 3d mapping to model the actual performance of pem (proton exchange membrane) fuel cell at various operating conditions," *Energy*, vol. 90, pp. 475–82, 2015.

[64] F. Al-Hadeethi, N. Haddad, A. Said, H. Alsyouri, and A. Abdelhadi, "Modeling hydrogen storage on mgeh2 and linh2 under variable temperature using multiple regression analysis with respect to anova," *International Journal of Hydrogen Energy*, vol. 42, no. 25558, p. e25564, 2017.

[65] M. Marciukaitis, I. Zutautaite, L. Martisauskas, B. Joksas, G. Gecevicius, and A. Sfetsos, "Non-linear regression model for wind turbine power curve," *Renewable Energy*, vol. 113, pp. 732–741, 2017.

[66] J. Rawlings, S. Pantula, and D. Dickey, *Applied regression analysis: a research tool*. Springer Science & Business Media, 2 ed., 2001.

[67] M. Srikanta and D. Akhil, "Chapter 4 regression modeling and analysis. applied statistical modeling and data analytics," *Applied statistical modeling and data analytics. A Practical Guide for the Petroleum Geosciences*, pp. 69–96, 2018.

[68] G. Seber and C. Wild, *Applied regression analysis: a research tool*. John Wiley & Sons, Hoboken, New Jersey, 2003.

[69] A. Ruckstuhl, "Introduction to nonlinear regression," *IDP Institut fur Datenanalyse und Prozessdesign, Zurcher Hochschule fur Angewandte Wissenschaften*, p. 365, 2010.

[70] F. Rudolf, J. William, and S. Ping, *Regression analysis: statistical modeling of a response variable*. Elsevier, 2 ed., 2006.

[71] A. Cameron and P. Trivedi, *Regression analysis of count data*. Cambridge University press, 2013.

[72] R. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," *In MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science, IEEE*, pp. 39–43, 1995.

[73] V. Ozsoy and H. Orkcu, "Estimating the parameters of nonlinear regression models through particle swarm optimization," *Gazi University Journal of Science*, vol. 29, no. 1, pp. 187–99, 2016.

[74] P. Erdogmus and S. Ekiz, "Nonlinear regression using particle swarm optimization and genetic algorithm," *International Journal of Computer Applications*, vol. 153, no. 6, pp. 28–36, 2016.

[75] S. Satapathy, J. Murthy, P. Reddy, B. Misra, P. Dash, and G. Panda, "Particle swarm optimized multiple regression linear model for data classification," *Applied Soft Computing*, vol. 9, no. 2, pp. 470–6, 2009.

[76] M. Hosseini, S. Naeini, A. Dehghani, and Y. Khaledian, "Estimation of soil mechanical resistance parameter by using particle swarm optimization, genetic algorithm and multiple regression methods," *Soil and Tillage Research*, vol. 157, pp. 32–42, 2016.

[77] Y. Jau, K. Su, C. Wu, and J. Jeng, "Modified quantum-behaved particle swarm optimization for parameters estimation of generalized nonlinear multiregressions model based on choquet integral with outliers," *Applied Mathematics and Computation*, vol. 221, pp. 282–95, 2013.

[78] R. Poli, J. Kennedy, and T. Blackwell, "Particle swarm optimization," *Swarm Intelligence*, vol. 1, pp. 33–57, 2017.

[79] Q. Bai, "Analysis of particle swarm optimization algorithm," *Computer and Information Science*, vol. 3, p. 180, 2010.

[80] M. Clerc, *Particle swarm optimization*. John Wiley & Sons, 2010.

[81] S. Talukder, "Mathematicle modelling and applications of particle swarm optimization," Master's thesis, School of Engineering at Blekinge Institute of Technology, 2011.

[82] M. Farman, Z. Iqbal, A. Ahmad, A. Raza, and E. Haq, "Numerical solution and analysis for acute and chronic hepatitis b," *International Journal of Analysis and Applications*, vol. 16, no. 6, pp. 842–55, 2018.

[83] E. Gulbandilar, A. Cimbiz, M. Sari, and H. Ozden, "Relationship between skin resistance level and static balance in type ii diabetic subjects," *Diabetes Research and Clinical Practice*, vol. 82, no. 3, pp. 335–339, 2008.

[84] D. Okuonghae, "A mathematical model of tuberculosis transmission with heterogeneity in disease susceptibility and progression under a treatment regime for infectious cases," *Applied Mathematical Modelling*, vol. 37, no. 10–11, pp. 6786–6808, 2013.

[85] S. Conoci, F. Rundo, S. Petralta, and S. Battiato, "In advanced skin lesion discrimination pipeline for early melanoma cancer diagnosis towards poc devices," *2017 European Conference on Circuit Theory and Design (ECCTD), IEEE*, pp. 1–4, 2017.

[86] M. Sari, "Relationship between physical factors and tibial motion in healthy subjects: 2d and 3d analyses," *Advances in Therapy*, vol. 24, no. 4, pp. 772–783, 2007.

[87] C. Liddell, N. Owusu-Brackett, and D. Wallace, "A mathematical model of sickle cell genome frequency in response to selective pressure from malaria," *Bulletin of Mathematical Biology*, vol. 76, no. 9, pp. 2292–2305, 2014.

[88] M. Sari and C. Tuna, "Prediction of pathological subjects using genetic algorithms," *Computational and Mathematical Methods in Medicine*, vol. 2018, 2018.

[89] J. McAllister, "Modeling and control of hemoglobin for anemia management in chronic kidney disease," 2017.

[90] A. Ngwira and L. Kazembe, "Analysis of severity of childhood anemia in malawi: a bayesian ordered categories model," *Open Access Medical Statistics*, vol. 6, pp. 9–20, 2016.

[91] V. Sharma and R. Kumar, "Dating of ballpoint pen writing inks via spectroscopic and multiple linear regression analysis: A novel approach," *Microchemical Journal*, vol. 134, pp. 104–113, 2017.

[92] X. Huang, Y. Yang, Y. Chen, C. Wu, R. Lin, Z. Wang, and X. Zhang, "Preoperative anemia or low hemoglobin predicts poor prognosis in gastric cancer patients: a meta-analysis," *Disease Markers*, vol. 2019, 2019.

[93] A. Saviano and F. Lourenco, "Measurement uncertainty estimation based on multiple regression analysis (mra) and monte carlo (mc) simulations–application to agar diffusion method," *Measurement*, vol. 115, pp. 269–278, 2018.

[94] J. Daru, J. Zamora, B. Fernández-Félix, J. Vogel, O. Oladapo, N. Morisaki, O. Tuncalp, M. Torloni, S. Mittal, and K. Jayaratne, "Risk of maternal mortality in women with severe anaemia during pregnancy and post partum: a multilevel analysis," *The Lancet Global Health*, vol. 6, no. 5, pp. e548–e554, 2018.

[95] P. Nguyen, S. Scott, R. Avula, L. Tran, and P. Menon, "Trends and drivers of change in the prevalence of anaemia among 1 million women and children in india, 2006 to 2016," *BMJ Global Health*, vol. 3, no. 5, p. e001010, 2018.

[96] A. Prieto, A. Silva, J. de Brito, J. Macías-Bernal, and F. Alejandre, "Multiple linear regression and fuzzy logic models applied to the functional service life prediction of cultural heritage," *Journal of Cultural Heritage*, vol. 27, pp. 20–35, 2017.

[97] M. Little, C. Zivot, S. Humphries, W. Dodd, K. Patel, and C. Dewey, "Burden and determinants of anemia in a rural population in south india: A cross-sectional study," *Anemia*, vol. 2018, 2018.

[98] N. Bessonov, A. Sequeira, S. Simakov, Y. Vassilevskii, and V. Volpert, "Methods of blood flow modelling," *Mathematical Modelling of Natural Phenomena.*, vol. 11, no. 1, pp. 1–25, 2016.

[99] M. Sari and A. Ahmad, "Anemia modelling using the multiple regression analysis," *International Journal of Analysis and Applications*, vol. 17, no. 5, pp. 838–849, 2019.

[100] A. Cetin and M. Sahin, "A monolithic fluid-structure interaction framework applied to red blood cells," *International Journal for Numerical Methods in Biomedical Engineering*, vol. 35, no. 2, p. e3171, 2019.

[101] H. Xu, Y. Mei, X. Han, J. Wei, P. Watton, W. Jia, A. Li, D. Chen, and J. Xiong, "Optimization schemes for endovascular repair with parallel technique based on hemodynamic analyses," *International Journal for Numerical Methods in Biomedical Engineering*, vol. 35, no. 6, p. e3197, 2019.

[102] S. Rivera and T. Ganz, "Animal models of anemia of inflammation," *Mediterranean Journal of Hematology and Infectious Diseases*, vol. 46, no. 4, pp. 351–357, 2009.

[103] K. Kawo, Z. Asfaw, and N. Yohannes, "Multilevel analysis of determinants of anemia prevalence among children aged 6—59 months in ethiopia: classical and bayesian approaches," *Anemia*, vol. 2018, 2018.

[104] D. Bates and D. Watts, *Nonlinear regression analysis and its applications*. John Wiley & Sons, New York, 1988.

[105] A. Ahmad and M. Sari, "Anemia prediction with multiple regression support in system medicinal internet of things," *Journal of Medical Imaging and Health Informatics*, vol. 10, no. 1, pp. 261–267, 2020.

[106] B. Cetiner and M. Sari, "ibial rotation assessment using artificial neural networks," *Mathematical and Computational Applications*, vol. 15, no. 1, pp. 34–44, 2010.

[107] Y. Loke, A. Harahsheh, A. Krieger, and L. Olivieri, "Usage of 3d models of tetralogy of fallot for medical education: impact on learning congenital heart disease," *BMC Medical Education*, vol. 17, no. 1, p. 54, 2017.

[108] M. Soler, M. Riera, and D. Batlle, "New experimental models of diabetic nephropathy in mice models of type 2 diabetes: efforts to replicate human nephropathy," *Experimental Diabetes Research*, vol. 2012, 2012.

[109] M. Stephens, A. Grey, J. Fernandez, R. Kalluru, K. Faasse, A. Horne, and K. Petrie, "3-d bone models to improve treatment initiation among patients with osteoporosis: A randomised controlled pilot trial," *Psychology & Health*, vol. 31, no. 4, pp. 487–97, 2016.

[110] D. Botesteanu, S. Lipkowitz, J. Lee, and D. Levy, "Mathematical models of breast and ovarian cancers," *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, vol. 8, no. 4, pp. 337–62, 2016.

[111] Q. Gilli, K. Mustapha, J. Frayret, N. Lahrichi, and E. Karimi, "Patient model for colon and colorectal cancer care trajectory simulation," *Health Science Journal*, vol. 11, no. 6, pp. 1–6, 2017.

[112] G. Barosi, M. Cazzola, S. M. M. Stefanelli, and S. Perugini, "Estimation of ferrokinetic parameters by a mathematical model in patients with primary acquired sideroblastic anaemia," *British Journal of Haematology*, vol. 39, no. 3, pp. 409–23, 1978.

[113] E. Mehrara, E. Forssell-Aronsson, V. Johanson, and L. K "A new method to estimate parameters of the growth model for metastatic tumours," *Theoretical Biology and Medical Modelling*, vol. 10, no. 1, p. 31, 2013.

[114] C. Berzuini, P. Franzone, M. Stefanelli, and C. Viganotti, "Iron kinetics: modelling and parameter estimation in normal and anemic states," *Computers and Biomedical Research*, vol. 11, no. 3, pp. 209–27, 1978.

[115] J. Kennedy and R. Eberhart, "Particle swarm optimization: Proceedings," *InIEEE international conference on neural networks*, vol. 4, pp. 1942–1948, 1995.

[116] Y. Shi and R. Eberhart, "A modified particle swarm optimizer," *In 1998 IEEE international conference on evolutionary computation proceedings. IEEE world congress on computational intelligence*, pp. 69–73, 1998.

[117] C. Yang and D. Simon, "A new particle swarm optimization technique," *In18th International Conference on Systems Engineering (ICSEng'05)*, pp. 164–169, 2005.

[118] F. Marini and B. Walczak, "Particle swarm optimization (pso). a tutorial," *Chemometrics and Intelligent Laboratory Systems*, vol. 149, pp. 153–65, 2015.

[119] S. Mohanty, "Particle swarm optimization and regression analysis–i," *Astronomical Review*, vol. 7, no. 2, pp. 29–35, 2012.

[120] K. Alzaidi, O. Bayat, and O. Ucan, "A heuristic approach for optimal planning and operation of distribution systems," *Journal of Optimization*, vol. 2018, 2018.

[121] A. Alfiyatin, R. Febrita, H. Taufiq, and W. Mahmudy, "Modeling house price prediction using regression analysis and particle swarm optimization," *International Journal of Advanced Computer Science and Applications*, 2017.

[122] H. Samareh, S. Khoshrou, K. Shahriar, M. Ebadzadeh, and M. Eslami, "Optimization of a nonlinear model for predicting the ground vibration using the combinational particle swarm optimization-genetic algorithm," *Journal of African Earth Sciences*, vol. 133, pp. 36–45, 2017.

[123] S. Chen, R. Yang, R. Yang, L. Yang, X. Yang, C. Xu, B. Xu, H. Zhang, Y. Lu, and W. Liu, "A parameter estimation method for nonlinear systems based on improved boundary chicken swarm optimization," *Discrete Dynamics in Nature and Society*, vol. 2016, 2016.

[124] H. Jahandideh and M. Namvar, "Use of pso in parameter estimation of robot dynamics; part two: robustness," *16th International Conference on System Theory, Control and Computing (ICSTCC), IEEE*, pp. 1–6, 2012.

[125] B. Choudhury and S. Neog, "Particle swarm optimization algorithm for integer factorization problem (ifp)," *International Journal of Computer Applications*, vol. 117, no. 13, 2015.

[126] A. Abdullah, S. Deris, M. Mohamad, and S. Anwar, "An improved swarm optimization for parameter estimation and biological model selection," *PloS one*, vol. 8, no. 4, p. e61258, 2013.

[127] H. Chu and L. Chang, "Applying particle swarm optimization to parameter estimation of the nonlinear muskingum model," *Journal of Hydrologic Engineering*, vol. 14, no. 9, pp. 1024–1027, 2009.

[128] H. Jahandideh and M. Namvar, "Use of pso in parameter estimation of robot dynamics; part one: No need for parameterization," *In System Theory, Control and Computing, ICSTCC, 2012, 16th International Conference, IEEE*, pp. 1–6, 2012.

Publications From the Thesis

Contact Information: arshed980@gmail.com

Papers

1. A. A. Ahmad and M. Sari, "Anemia Prediction with Multiple Regression Support in System Medicinal Internet of Things", Journal of Medical Imaging and Health Informatics, Vol. 10, No. 1, 261–267, 2020.

2. M. Sari and A. A. Ahmad, "Anemia Modelling Using the Multiple Regression Analysis", International Journal of Analysis and Applications, Vol. 17, No. 5, 838-849, 2019.

3. A. A. Ahmad and M. Sari, "Parameter Estimation to an Anemia Model Using the Particle Swarm Optimization", Sigma Journal of Engineering and Natural Sciences, Vol. 37, No. 4, 1331-1343, 2019.

4. A. A. Ahmad, M. Sari and T. Coşgun, "A Medical Modelling Using Multiple Linear Regression", In Mathematical Modelling and Optimization of Engineering Problems, Springer, Cham, Vol. 30, 71-87, 2020.

5. M. S. Mohammed, A. A. Ahmad and M. Sari, "Analysis of Anemia Using Data Mining Techniques with Risk Factors Specification", Accepted in (INCET2020) IEEE International Conference for Emerging in Technology, Belgaum, India 2020.

6. M. Sari and A. A. Ahmad, "Predicting Anemia in Medical Systems Using Artificial Neural Networks", 2019. (Submitted)

7. M. Sari, A. A. Ahmad and H. Uslu, "Medical Model Estimation with Particle Swarm Optimization", 2019. (Submitted)

8. M. Sari, A. A. Ahmad and H. Uslu, "Advection-Diffusion Process through a Difference Scheme-Based Monte Carlo Simulation", 2019. (Submitted)

9. A. A. Ahmad, M. Sari and A. A. Ahmed, "An Efficient Algorithm to Predict Anemia in an Educational Way", 2019. (Submitted)

10. M. Sari, M. S. Mohammed, A. A. Ahmad and I. Dag, "Anemia Prediction using Long-Short Term Memory Technique", 2020. (Submitted)

11. A. A. Ahmad, M. Sari and I. Demir, "Anemia Prediction Model Through Path Analysis Approach", 2020. (Submitted)

12. M. Sari, E. Kasap, A. A. Ahmad and H. Uslu, "Prediction of Hepatitis B Immunization Using Genetic Algorithm", 2020. (Submitted)

Conference Papers

1. M. Sari, A. A. Ahmad and M. Kirisci, Use of the Monte Carlo Methods, Computational Sciences Research, Istanbul-Turkey, 2016.

2. M. Sari, L.J.M. Al-Mashhadani and A. A. Ahmad, Discussion on Advection-Diffusion Equation Through Finite Difference Schemes, 2^{nd} International Conference on Analysis and Its Applications, Kirşehir-Turkey, 2016.

3. M. Sari, L.J.M. Al-Mashhadani and A. A. Ahmad, Numerical analysis of convection–diffusion processes, International Congress on Fundamental and Applied Sciences, Istanbul-Turkey, 2016.

4. M. Sari, A. A. Ahmad and L. Al-mashhadani, Capturing the Behavior of Advection-Diffusion Process Through Monte Carlo Simulation , 2^{nd} International Conference on Computational Mathematics and Engineering Sciences (CMES-2017), Istanbul, 2017.

5. M. Sari, A. A. Ahmad and T. Coşgun, Mathematical Modelling of a Medical Problem Using Multiple Linear Regression, International Conference on Applied Mathematics in Engineering (ICAME), Balikesir, Turkey, 2018.

6. T. Coşgun, M. Sari and A. A. Ahmad, A Discussion of the Singular and Weakly Singular Integral Equations with Abel Type Kernels, International Conference on Applied Mathematics in Engineering (ICAME), Balikesir, Turkey, 2018.

7. M. Sari, E. Kasap and A. A. Ahmad, Prediction of Hepatitis B Immunization Using Genetic Algorithm, 4^{th} International Conference on Computational Mathematics and Engineering Sciences, (CMES 2019), Antalya, Turkey, 2019.

8. M. S. Mohammed, A. A. Ahmad and M. Sari, Analysis of Anemia Using Data Mining Techniques with Risk Factors Specification, (INCET2020) IEEE International Conference for Emerging in Technology, Belgaum, India 2020.