DEVELOPMENT OF ARTIFICIAL NEURAL NETWORK MODELS FOR PREDICTING LIPID PROFILE USING SMARTMF ELECTRICAL PARAMETERS

AHMAD ZULKHAIRI BIN ZULKEFLI

FACULTY OF ENGINEERING
UNIVERSITY OF MALAYA
KUALA LUMPUR

2021

DEVELOPMENT OF ARTIFICIAL NEURAL NEWORK MODELS FOR PREDICTING LIPID PROFILE USING SMARTMF ELECTRICAL PARAMETERS

AHMAD ZULKHAIRI BIN ZULKEFLI

RESEARCH PROJECT SUBMITTED TO THE FACULTY OF ENGINEERING UNIVERSITY MALAYA, IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF BIOMEDICAL ENGINEERING

FACULTY OF ENGINEERING
UNIVERSITY OF MALAYA
KUALA LUMPUR

2021

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Ahmad Zulkhairi Bin Zulkefli

Matric No: KQB190002 @ 17201268/1

Name of Degree: Master of Biomedical Engineering

Title of Research Report:

"Development of Artificial Neural Network for Predicting Lipid Profile Using SmartMF Electrical Parameters"

Field of Study: Biomedical Engineering

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date: 8 Feb 2021

Subscribed and solemnly declared before,

Witness's Signature Date:

Name:

Designation:

DEVELOPMENT OF ARTIFICIAL NEURAL NETWORK MODELS FOR PREDICTING LIPID PROFILE USING SMARTMF ELECTRICAL PARAMETERS

ABSTRACT

This research project aims in the development of artificial neural network (ANN) to classify lipid profile parameters; total cholesterol (TC), triglycerides (TG), high density lipoprotein cholesterol (HDL-C) and low-density lipoprotein cholesterol (LDL-C) by utilizing a non-invasive bioelectrical impedance analysis (BIA) device (SmartMF). Data were obtained from volunteer patients requiring lipid profile testing at the primary care department of University Malaya Medical Centre (UMMC).

The data were divided into two categories for each lipid profile parameters, normal and abnormal. Statistical analysis including T-Test and logistic regression analysis for the bioelectrical parameters were employed to understand the data and determine suitable predictors to be used for the development of the ANN. Results of the statistical analysis indicates capacitance, resistance and reactance as significant predictor for TC level. Impedance at 50, 100 and 200 kHz, resistance and reactance as significant predictors for TG level. Impedance at 5, 50, 100 and 200 kHz as significant predictors for HDL-C level. No significant predictors were determined for LDL-C level, thus ANN model for the parameter cannot be developed.

ANN employing the multi-layered feed forward neural network technique was developed for the TC, TG and HDL-C parameters utilizing the scaled conjugate gradient (SCG), Levenberg Marquardt (LM) and Resilient (RB) backpropagation algorithm. The best model was then selected based on the testing performance.

For TC, the SCG model with a testing accuracy of 73.3%, sensitivity of 63.6% and specificity of 78.9% was selected. For TG, the LM algorithm with a testing accuracy of 76.7%, sensitivity of 28.6% and specificity of 91.3% was selected. For HDL-C, the SCG

algorithm with a testing accuracy of 76.7%, sensitivity of 95.5% and specificity of 25.0% was selected as best performing models to indicate the relationship of the bioelectrical parameters and the lipid profile parameters.

Keywords: artificial neural network (ANN), bioelectrical impedance analysis (BIA), lipid profile, SmartMF

PEMBANGUNAN RANGKAIAN NEURAL BUATAN BAGI MERAMAL PROFIL LIPID MENGGUNAKAN PARAMETER ELEKTRIK SMARTMF

ABSTRAK

Kajian ini bertujuan untuk membangunkan sebuah rangkaian neural buatan (ANN) untuk mengklasifikasikan parameter profil lipid; jumlah cholesterol (TC), trigliserida (TG), kolesterol lipoprotein berketumpatan tinggi (HDL-C) dan kolesterol lipoprotein berketumpatan rendah (LDL-C) menggunakan pendekatan tidak invasif dengan peranti analisis galangan bioelektrik (BIA) (Ukuran Komposisi Badan SmartMF). Data diperolehi daripada sukarelawan di kalangan pesakit yang memerlukan ujian profil lipid di jabatan penjagaan kesihatan primer University Malaya Medical Centre (UMMC).

Data dibahagikan kepada dua kategori untuk setiap parameter profil lipid, normal dan tidak normal. Analisis statistik iaitu T-test dan analisis regresi logistik bagi parameter bioelektrik digunapakai untuk pemahaman data bagi menentukan peramal yang sesuai digunakan untuk pembangunan ANN. Keputusan analisis statistik tersebut menunjukkan kapasitans, rintangan dan reaktans sebagai peramal yang signifikan bagi paras jumlah kolesterol (TC). Galangan di paras 50, 100 dan 200 kHz, rintangan dan reaktans merupakan peramal yang signifikan bagi paras TG. Galangan di paras 5, 50, 100 dan 200 kHz merupakan peramal yang signifikan bagi paras HDL-C. Tiada peramal yang signifikan dapat ditentukan bagi paras LDL-C, oleh itu model ANN bagi parameter tersebut tidak dapat dibangunkan.

ANN yang menggunapakai teknik rangkaian neural suap-hadapan pelbagai lapisan telah dibangunkan bagi parameter TC, TG dan HDL-C menggunakan algoritma gradien konjugasi berskala (SCG), Levenberg Marquardt (LM) dan *resilient backpropagation* (RB). Model yang terbaik kemudiannya dipilih berdasarkan prestasi ujian tertinggi.

Algoritma SCG dipilih bagi parameter TC dengan ketepatan ujian 73.3%, sensitivity ujian 63.6% dan spesifisiti ujian 78.9%. Bagi TG, algoritma LM dengan ketepatan ujian

76.7%, sensitivity ujian 28.6% dan spesifisiti ujian 91.3% telah dipilih. Bagi HDL-C, algoritma SCG dengan ketepatan ujian 76.7%, sensitivity ujian 95.5% dan spesifisiti ujian 25.0% telah dipilih sebagai model berprestasi terbaik bagi menunjukkan hubungan antara parameter bioelektrik dan parameter profil lipid.

Kata kunci: rangkaian neural buatan (ANN), analisa impedans bioelektrikal (BIA), profil lipid, ukuran komposisi badan SmartMF

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincerest gratitude and acknowledgements to the following

My supervisor Ir. Dr Mas Sahidayana Mohktar for her trust, patience and support for me to undertake this research project. Her guidance throughout the research had helped me gained valuable knowledge and experiences for my future undertakings.

My supervisor from University Malaya Medical Centre, Prof. Dr. Sajaratulnisah Othman for accommodating me in the centre for data collection.

My fellow research assistant Mohd Iz'aan Paiz Zamri, who had assisted and imparted invaluable experience working in the field.

My family especially my parents Zulkefli Mohd Ghazali and Norhasimah Othman for their lasting support, care and love for me to go through any ordeal during this research

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

ACC	: American College of Cardiology
AHA	: American Heart Association
ANN	: Artificial neural network
ART	: adaptive resonance theory
BCM	: body cell mass
BIA	: bioelectrical impedance analysis
BIS	: bioelectrical spectroscopy
BIVA	: bioelectrical impedance vector analysis
BMI	: body mass index
CHD	: coronary heart disease
CV	: cardiovascular
CVD	: cardiovascular disease
ECW	: extracellular water
FFM	: fat-free mass
FN	: false negative
FP	: false positive
FRS	: Framingham Risk Score
HDL-C	: high density lipoprotein cholesterol
ICW	: intracellular water
LDL-C	: low density lipoprotein cholesterol
LM	: Levenberg Marquardt
LP	: lipid profile
LR	: logistic regression
MF	: multi frequency

MFNN : multi-layered Feed forward neural network

MLP : multi-layered perceptron

NALFD : non-alcoholic fatty liver disease

NHMS : National Health and Morbidity Malaysia

NN : neural network

PSPD : Pusat Setempat Pengambilan Darah

QDA : quadratic discriminant analysis

RB : resilient backpropagation

SCG : scaled conjugate gradient

SD : standard deviation

SF : single frequency

SOM : self-organizing map

TBW : total body water

TC : total cholesterol

TG : triglyceride

TN : total negative

TP : total positive

UM : University Malaya

UMMC : University Malaya Medical Centre

VLDL-C : very low-density lipoprotein cholesterol

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

1.1 Overview

A lipid profile is a common blood test performed in the clinical healthcare services as part of a routine tests to predict the general well-being of patients especially in the outcome of cardiovascular diseases. The basis of the test is the measurements of varying types of cholesterol components in the human body. Cholesterol is an essential substance for normal physiological functions through various means such as cellular functions, hormone production and a multitude of biochemical processes in the body. However, any abnormalities in its levels could indicate homeostatic dysfunctions caused by certain conditions or diseases that warrants further investigations.

Venepuncture is the standard procedure required to procure the blood samples for lipid profile testing. This procedure is invasive and carries its own sets of risks in the form of infection, formation of hematoma, bruises and most importantly causes discomfort towards patient especially in cases where multiple puncturing is needed to obtain the necessary samples (Galena, 1992). Thus, there is a need to explore the possibility of non-invasive solutions to determine the lipid profile to mitigate those issues.

Bioelectrical Impedance Analysis (BIA) have the potential of providing a non-invasive modality to predict the cholesterol levels of the lipid profile. BIA has been a popular alternative in the assessment of human body compositions through the provision of parameters that are indicative of the health status of an individual. Therefore, in this study, the potential use of BIA parameters in classifying the cholesterol levels will be investigated through the application of an Artificial Neural Network (ANN) in providing the needed association between the variables and thus provides a classification model for the lipid profile parameters. This will enable the possibility of utilizing BIA devices as a screening tool for cholesterol levels.

A review has shown that ANN has been used to produce classification of data in the healthcare field for a number of purposes including risk assessment and diagnostics (Shahid, Rappon, & Berta, 2019).

1.2 Problem Statement

Standard procedure to measure the lipid profile of a patient requires an invasive procedure to obtain the blood samples for laboratory investigation and could impede better detection of dyslipidaemia among the population. Is there a possible non-invasive approach to screen for abnormal lipid profile measurements?

1.3 Objectives

The objectives of this research project are as following

1. Determination of significant predictors from SMARTMF electrical parameters for each measurement in the lipid profile
2. Development of an Artificial Neural Network classification models using the SMARTMF electrical parameters to understand its association with the lipid profile of a subject

1.4 Scope of work

The scope of the research includes data collection of patients from the Department of Primary Care Medicine in University Malaya Medical Centre. Significant predictors are then determined for the cholesterol levels of the lipid profile based on the SMARTMF electrical parameters through statistical analysis. Classification model will be developed using ANN for the parameters of the lipid profile utilizing chosen predictors from the SMARTMF electrical parameters.

CHAPTER 2: LITERATURE REVIEW

2.1 Lipid Profile

2.1.1 Constituents of Lipid Profile

The constituent of a lipid profile comprises of four basic parameters; the total cholesterol (TC), high density lipoprotein cholesterol (HDL-C), low density lipoprotein cholesterol (LDL-C) and triglycerides (TG). Another component in the lipid profile is very low-density lipoprotein cholesterol (VLDL-C). VLDL-C in not commonly included in a lipid profile as there is no simple way of measuring it. An estimation is usually performed through the TG level but is not applicable when the TG level is too high.

The total cholesterol is an estimate of the overall cholesterol level which includes the HDL-C, LDL-C and VLDL-C. The HDL-C is considered as the 'good' cholesterol as it functions to take up cholesterols from the cardiovascular system to the liver to be eliminated from the body. Meanwhile the LDL-C and VLDL-C is considered as 'bad' cholesterol as it serves to transport lipid from the liver to the cardiovascular system and risks deposition of the substances on the vessel walls. On the other hand, TG is a type of fat (lipid) that are different from cholesterol. It serves to store excess calories in the liver and when needed will release the calories to provide energy for the body. Thus, TG is also used as an indicator of metabolic dysfunction and cardiovascular disease. Table 2.1 indicates the normal values for the parameter of a standard lipid profile in clinical practice (MIMS, 2020).

Table 2.1: Dyslipidaemia Lipid Profile Range

Parameters	Lipid Level
TC	> 5.2 mmol/L
HDL-C	< 1.6
LDL-C	> 2.6 mmol/L
TG	> 1.7 mmol/L

Standard diagnostic test for lipid profile utilizes enzymatic analysis performed in laboratories that requires blood samples through invasive means. The accuracy of the analysis is very high with an error rate of less than 2% (Warnick, Kimberly, Waymack, Leary, & Myers, 2008).

2.1.2 Indications and Clinical Significance

There are several indications for lipid profile testing based on the Malaysian Clinical Practice Guideline for Management of Dyslipidaemia 2017. This include screening for all adults > 30years of age, the presence of other cardiovascular risk factors, individuals with known high risk of developing CVD earlier in life as in cases of those with a family history of premature CVD, metabolic syndrome, diabetes and genetic dyslipidaemias (Ministry of Health Malaysia, 2017).

Dyslipidaemia is defined as abnormal levels of one or more kinds of lipid (fat) in the blood system. It is one of the conditions listed under the umbrella term of metabolic syndrome. If an individual is deemed to have dyslipidaemia, it means that either the level of LDL-C or TG are too high or the HDL-C level is too low (Fodor, 2008). The abnormality in the level of cholesterol in a patient can cause pathologies through cholesterol plaque formation known as atherosclerosis. Dyslipidaemia is widely

considered as a major prerequisite to the development of atherosclerotic cardiovascular disease (CVD) (Kopin & Lowenstein, 2017; Urbina & Daniels, 2008).

Atherosclerosis is a complex process characterized by the build-up of lipids, cholesterol, cellular debris and calcium, inflammation and dysfunction of the endothelial tissues in the vessel wall (Hansson, 2005). As the atherosclerotic plaque grows, patency of the vessels declines reducing supply of blood flow to organs (Stoll & Bendszus, 2006). As this worsens, organs will undergo ischaemia and eventually infarction. Atherosclerotic CVD can generally be classified into coronary heart disease, cerebrovascular heart disease and peripheral arterial disease. Common conditions of concern in CVD are heart attacks or myocardial infarction, stroke and limb ischaemia (World Health Organization, 2017).

For a decade, CVD has been the primary cause of morbidity and mortality in Malaysia and the main cause of global mortality (World Health Organization, 2018). A study for the National Health and Morbidity Malaysia (NHMS) 2015 has shown the prevalence of hypercholesterolemia which is a type of dyslipidaemia in Malaysian population is at 47.7% and it is also a concern that only 19.2% of them are actually aware of their abnormal cholesterol status (Mat Rifin et al., 2018). Other risk factors for CVD includes age, gender, ethnicity, occupation, education level and other diseases categorized under metabolic syndrome; obesity, hypertension and diabetes (National Health Services UK, 2018).

2.2 Bioelectrical Impedance Analysis (BIA)

2.2.1 Working Principle of BIA

Bioelectrical Impedance Analysis (BIA) is a technique based on the principle that the resistance (R) or impedance (Z) value of a particular material with uniform cross-sectional area and homogenous conductive properties is proportional to its length (L) and

inversely proportional to its cross-sectional area (A) as shown in Figure 2.1. It is to note that the human body does not have constant conductivity and is not a uniform cylinder. It is suggested that an empirical relationship can be established between the impedance quotient (L^2/R) and the volume of water in the body that composed of conducting electrolytes. The length parameter is equal to the height of the human body usually measured from wrist to ankle. Hence an empirical correlation between body mass (consist of 73% water) and height^2/R is established. The human body however is not an equivalent cylinder thus there is a need to match it with real geometry by a suitable coefficient, ρ which is the resistivity that is dependent on a number of factors (Kyle, Bosaeus, De Lorenzo, Deurenberg, Elia, Gómez, et al., 2004) .

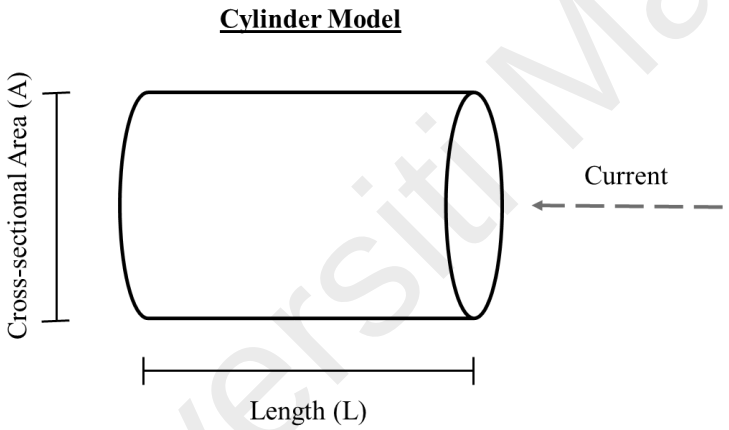

Figure 2.1: Cylinder model depicting relationship of impedance and geometry. Resistance (R) = ρL/A = ρL^2/V. V equals to AL and represents the resistivity of the conducting material.

The human body is complex as the resistance to an electrical current exist in two forms, the reactance which is the capacitive R and the resistance which is the resistive R. The cell membranes contribute to the capacitance value while the intra and extracellular fluids contributes to the resistive value. The combination of these two parameters is called impedance. Many studies describes electrical behaviours of human tissues through representation of electrical circuits (Gudivaka, Schoeller, Kushner, & Bolt, 1999). Figure 2.2, shows a common electrical circuit representation of the human body. At low

frequency (ideally zero), the electrical current does not enter the cell membrane. Thus, current only passes through the extracellular fluid and impedance is measured as R0. In opposition, at very high frequency (ideally infinite) the cell membrane acts as a perfect capacitor and hence allowing current to pass through and impedance value will be based on the total R of both intra and extracellular fluid (R∞). A Cole-Cole plot represent the relationship of the parameters as the use of zero and infinite frequency is not practical (Figure 2.3) (Cole & Cole, 1941).

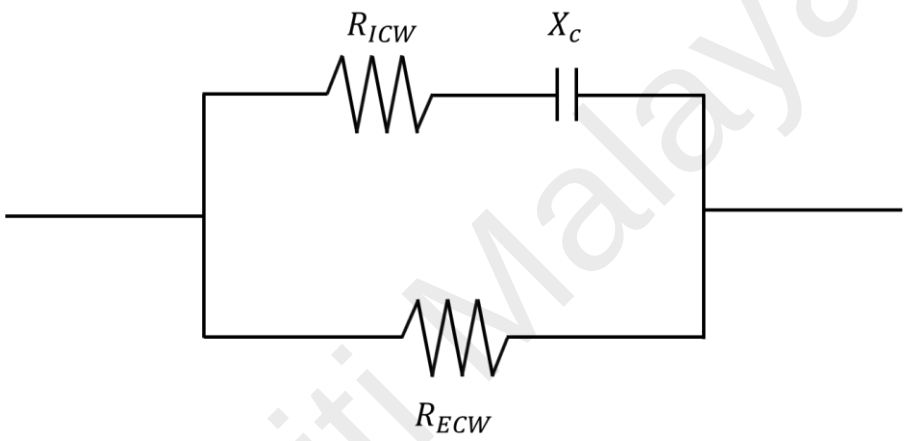

Figure 2.2: Simple electrical circuit model of a human body. X_c = Capacitance from cell membrane and R = Resistance of intra and extracellular fluids.

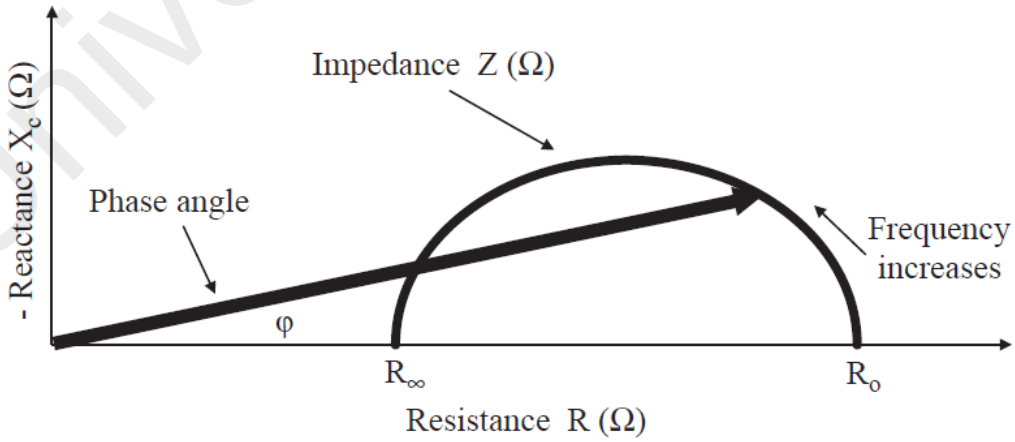

Figure 2.3: Graphical representation of phase angle and its relationship with Resistance (R), Reactance (X_c), Impedance (Z) and frequency of applied current (Kyle, Bosaeus, De Lorenzo, Deurenberg, Elia, Gómez, et al., 2004).

The relationship of the parameters in Figure 2.3 reflects the different electrical properties of tissues predisposed by many factors such as disease pathology, hydration status and nutritional status. This relationship measured by the phase angle and other interrelated parameters are utilized to predict clinical outcome (Schwenk et al., 2000; Toso et al., 2000).

2.2.2 Classification and Parameters of BIA

BIA is classified based on the method of analysis. There are 5 types of BIA currently available; single frequency BIA (SF-BIA), multi frequency BIA (MF-BIA), Bioelectrical spectroscopy (BIS), segmental BIA and Bioelectrical Impedance Vector Analysis (BIVA) (Kyle, Bosaeus, De Lorenzo, Deurenberg, Elia, Gómez, et al., 2004). SF-BIA works at the frequency of 50 kHz while MF-BIA works at multiple frequencies from (0 – 500 kHz) and both employs empirical linear regression models and mixture modelling to estimate the parameters. MF-BIA was shown to be more precise in the assessment of extracellular water (ECW) however it is less accurate in the assessment of total body water (TBW) compared to SF-BIA (Patel, Peterson, Silverman, & Zarowitz, 1996). Compared to the first two types of BIA, BIS utilizes mathematical modelling to establish the relationship between impedance and the body fluid compartments (Cornish, Ward, Thomas, Jebb, & Elia, 1996). BIVA on the other hand does not depend on any equations or models to generate relationship instead relies on direct measurement of the impedance vector to evaluate patient. Body composition serves as the parameters which includes fat free mass (FFM), extracellular water (ECW) and intracellular water (ICW) that makes up the total body water (TBW) and body cell mass (BCM) (Figure 2.4). These parameters are derived or estimated through the modelling of electrical parameters i.e., impedance, body capacitance, resistance reactance and phase angle which are the main outputs of bioelectrical impedance analysis devices.

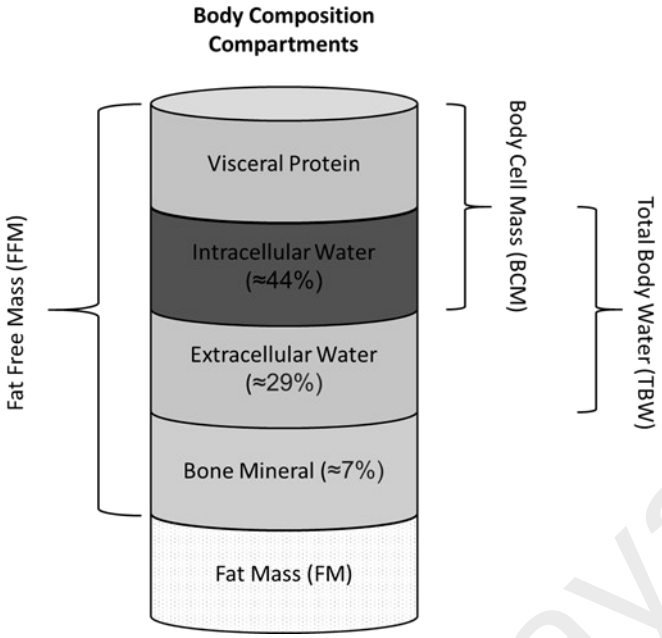

Figure 2.4: Compartments of a human body

2.2.3 SmartMF BIA Body Composition Measurement Device

The SmartMF is a type of multifrequency bioelectrical impedance analysis (MF-BIA) device that functions to measure the electrical properties of a body, then used to derive the body composition parameters. It is a portable device with a software module enabling results to be shown on a smartphone through its dedicated application (Figure 2.5). The system was developed and validated by biomedical engineers and healthcare professionals from University Malaya. SmartMF has received a number of national and international achievements and awards for its innovation and its ability to assess total cholesterol level in subjects.

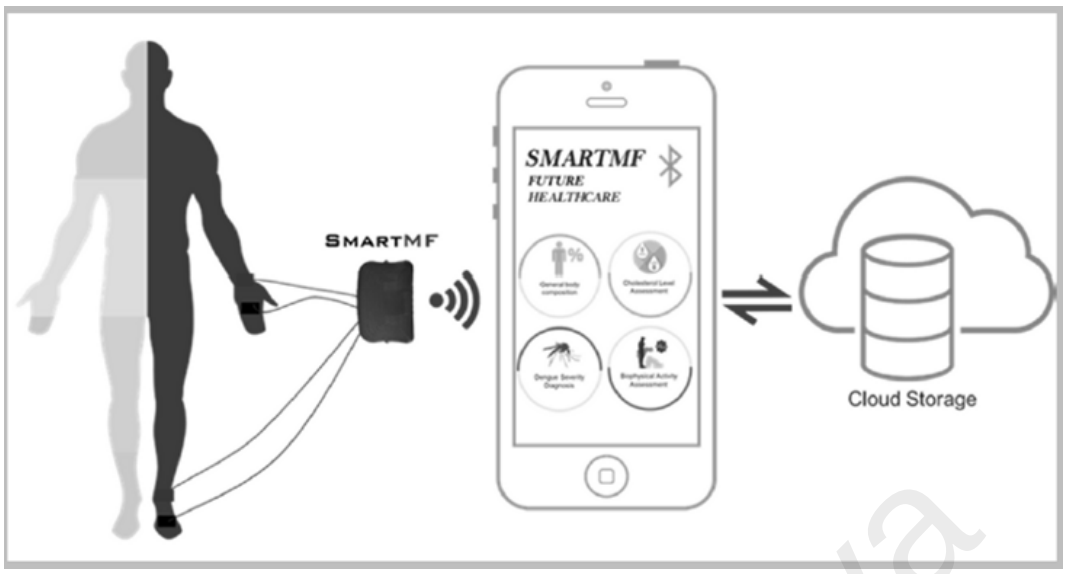

Figure 2.5: SmartMF Body Composition Measurement System

The device comprises of placement of four electrodes – two on the right foot and ankle and another two on the right hand and wrist. The electrodes are connected to a small electrical device. A minute electrical signal of (<1mA) current and of multiple range of frequencies (5-200kHz) are passed through from a set of electrodes through the subject's body into another set of electrodes. The basic principle of the BIA system is that differing body composition such as body weight, muscle mass and fat mass produces different electrical impedance. This will then be reflected on the parameter produced by the device. SmartMF system is still under further development and requires extensive validation for it to be implemented in clinical practice. One of the methods to validate such devices is through correlation of its parameter with the actual biochemical level of a subject in this case the lipid profile.

2.2.4 Existing Studies on BIA in the Clinical Field

Bioelectrical Impedance (BIA) has been studied considerably in the field of medicine especially on its applications in diseases affecting body composition parameters and researches. Kyle U.G. et. al, 2004 reviewed the application of BIA in the clinical field and reported it as a promising technique in the assessment and monitoring of health status

in healthy subjects and those with chronic diseases due to its characteristics of being non-invasive, inexpensive, good safety and portability. However, further evaluations must be made in cases of extreme ranges of BMI and abnormal hydration status (Kyle, Bosaeus, De Lorenzo, Deurenberg, Elia, Manuel Gómez, et al., 2004). This section describes some of the studies involving the utilization of BIA in the clinical field.

A recent multicentred clinical study evaluates the significance of segmental phase angle (PhA) parameters from a bioelectrical impedance analysis in the purpose of screening and monitoring of disease progression in patients with diabetes mellitus. A reduction in the PhA values were observed in diabetic patient as compared to non-diabetic patient ($p<0.001$). The reduction is more evident with the increase in the duration of diabetic morbidity (Jun, Ku, Kim, Kim, & Kim, 2020).

Kwon Y. and Jeong S.J., 2020, in their study utilizes the skeletal muscle mass to body fat ratio (MFR) derived from MF-BIA as a predictor for non-alcoholic fatty liver disease (NALFD) in non-obese children and adolescent. It was concluded that the association of NALFD and low MFR in the target population to be significant ($p=0.016$), thus providing a possible management strategy for the condition (Kwon & Jeong, 2020).

A cross sectional study in Brazil, employs SF-BIA to derive phase angle (PhA) parameters and investigate its correlation to first cardiovascular (CV) event risk based on the scores of American College of Cardiology/American Heart Association (ACC/AHA) and the Framingham Risk Score (FRS). Outcome from the study reveals that higher PhA is associated with lower risk of CV event for patients in high-risk group. Therefore, PhA can be a possible tool for primary prevention of CV events (Portugal et al., 2020).

2.3 Artificial Neural Network

2.3.1 Overview of Artificial Neural Network

Artificial Neural Network (ANN) takes inspiration from the inner working of a human brain by simulating the network of neurons making computers capable of producing decisions in a humanlike manner (Walczak & Cerpa, 2003). The simulating algorithm allows the network to 'learn' ways to solve a multitude of real-life problems (Krogh, 2008). Figure 2.6 shows a sample of ANN and its interconnected network correlating the inputs with the outputs. The neurons are arranged in a layer with the output of one layer serving as the input to the next layer. Each individual neuron is either connected to all neurons of the succeeding layer or a subset of it creating the synaptic connection of the brain. The input values of a processing element (in) are multiplied by the weight of the connection (wn) replicating the strengthening of the neural pathways. Learning in ANN is emulated through the adjustment of the connection strength or weight (Walczak & Cerpa, 2003).

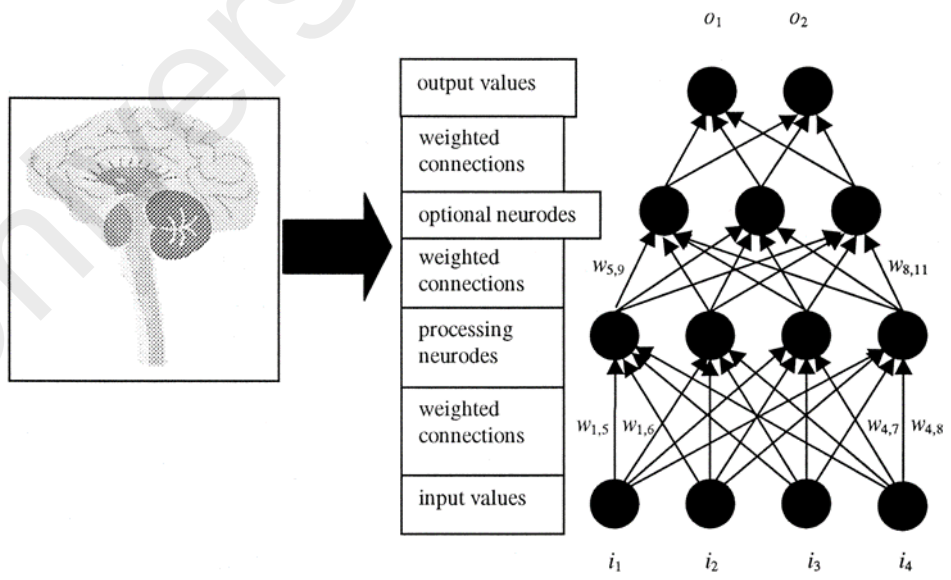

Figure 2.6: Sample of Artificial Neural Network and its components (Walczak & Cerpa, 2003)

2.3.2 Supervised Learning Algorithm

Supervised learning of ANN requires input or augmentations from the teacher to train said network based on their knowledge through labelled data sets (Fabiyi, 2019).

The multilayer perceptron (MLP) with a backpropagation algorithm is based on a supervised procedure that involves building a model through examples of input data with known output. This produces a general correlation between a particular problem and a solution (Lek & Park, 2008).

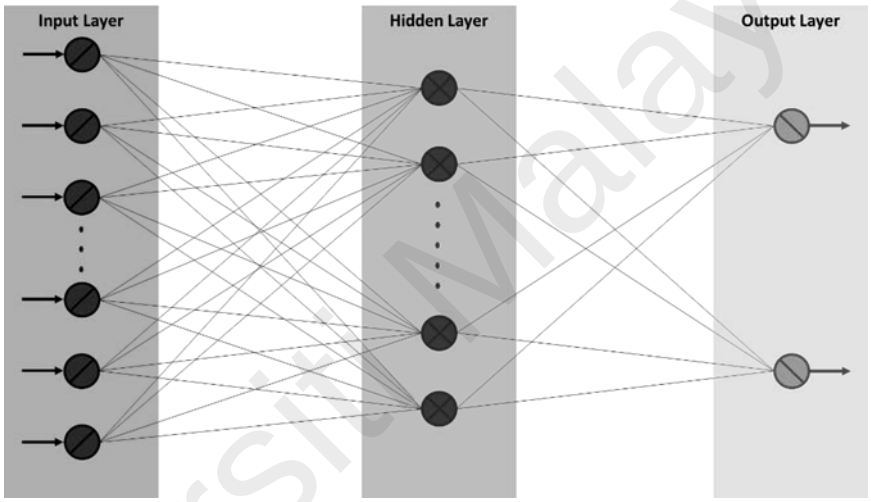

Figure 2.7: Multi-layered feed forward neural network (MFNN)/ MLP Architecture

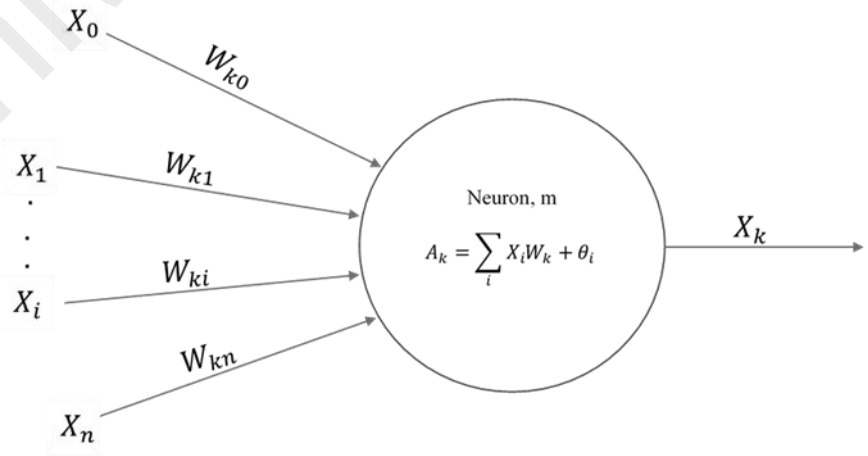

Figure 2.8: Forward propagation process. Input value (X_i), Weight (W_{ki}), Activation (k), Offset/Bias (θ_i) and Output Value (X_k)

The structure of an MLP is based on a layered feed-forward neural network in which the neurons are organised in sequential layers and the data flows in one direction from input layer to output layer through a set of hidden layers shown in Figure 2.7. During the forward-propagating step each connection from the input to neuron is associated with a weight or connection strength (Figure 2.8). A net input (called activation) is calculated which is the sum of each neuron with it associated weights expressed as in Equation 2.1

$$A_k = \sum_i X_i W_{ki} + \theta_i \quad \ldots\ldots \text{Equation 2.1}$$

Once calculated, the output value can be determined by applying a transfer function (Equation 2.2),

$$X_k = f(A_k) \ldots\ldots \text{Equation 2.2}$$

Based on the relationship between the input and output, different transfer function can be applied such as sigmoid function, threshold function, linear function etc.

The backward-propagating step of the MLP follows after the forward-propagation. This step determines the error through comparison of the network's output and the targeted output value during the training process of the network. This error is then utilized to modify the weight for the next sets of data to ensure the errors are at the minimum level possible. There are multiple backpropagation methods available including the Levenberg-Marquardt, Resilient and the Scaled Conjugate Gradient.

In the testing phase, the testing data are fed to the neural network. The output given by the neural network is compared with the desired output. The agreement or the disagreement of these two sets of output indicates the performance of the established neural network model. A new third independent data should then be utilized to validate the trained network. (Lek & Park, 2008)

2.3.2.1 Scaled Conjugate Gradient Backpropagation Algorithm

A conjugate gradient algorithm utilizes second order techniques rather than first order technique seen in standard backpropagation which performs better in improving the weight to the minimum level. The Scaled Conjugate Gradient (SCG) algorithm is a complex algorithm that are design to circumvent the line search-based algorithms employed by other conjugate gradient techniques thus improving computing time (Møller, 1993).

2.3.2.2 Levenberg-Marquardt Backpropagation Algorithm

The Levenberg-Marquardt (LM) algorithm is an algorithm utilized to solve non-linear least square problems through curve fitting. Its curve fitting is based on the combination of the gradient descent and the Gauss-Newton. The algorithm updates each iteration or weight through either of the method (Gavin, 2013). The advantage of the LM algorithm is that it is robust due to the two different method within (Nelles, 2013). However it can be slow in converging when dealing with model using more than 10 parameters (Gutenkunst et al., 2007).

2.3.2.3 Resilient Backpropagation Algorithm

In general, multi-layered networks employs sigmoid transfer function in the hidden layers. Training a network using the function only causes small changes in the weight after each iteration due to the small magnitude of the gradient although the weight is far from optimal (Gupta & Kang, 2011). The Resilient Backpropagation (RB) algorithm is a first order algorithm that aims to eliminated the effect of the small gradient magnitude and augments the size of the weight purely determined by a weight-specific update value. (Mohammed, Haithem, Alani, & George, 2015)

2.3.3 Unsupervised Learning

Unsupervised learning as the name implies does not involve a teacher who uses his or her knowledge to train the network. It is mostly used in cases where there are no possibilities to augment the network data sets with class identifiers. These usually occur when there is no availability of knowledge of the data environments or the cost of procuring it is too high (Atiya, 1990). Instead of relying in data augmentations, unsupervised learning networks are supplied with unlabelled input only data set and is required to discover patterns from it to build a newer model. This is achieved through the categorization based on the distance between clusters of data within (Fabiyi, 2019).

Examples of unsupervised learning algorithm includes the Hopfield networks, adaptive resonance theory (ART) and the most basic which is the self-organizing map (SOM) (Walczak & Cerpa, 2003).

2.3.4 Previous Studies on Artificial Neural Network in the Clinical Field

Development and advancement in the health care systems has revolutionize health care delivery model into a value based and patient centred model of care. This transformation however introduces new complexities in its implementation and integration into the health care system (Kuziemsky, 2016). Artificial intelligence provides a possible solution to this issue. A study of artificial neural network reveals that it can be implemented across all levels of health care decision making in its ability to be personalized to any specific needs allowing a much more effective and efficient delivery of health care (Shahid et al., 2019). This section will describe studies involving ANN in the clinical field that are available or under development. There are no recent studies that utilizes ANN in developing a classification model through data obtain from BIA.

A recent study utilizes artificial neural network for the prediction of lipid profile. The study uses measures of obesity and anthropometry such as Age, BMI, waist

circumference and abdominal diameter to develop the neural network model. The models produces reliable accuracy for TC (81.89%), LDL (79.29%), and HDL (79.29%) but poor predictive accuracy for TG (44.48%) (Vrbaški, Doroslovački, Kupusinac, Edita, & Ivetic, 2019)

A study has been conducted in the implementation of ANN model for prediction of risk and major complication for diabetic patients as standard models that are available have the limitations of relying on simple statistical methodology and only covers major complications such as cardiovascular disease diabetic retinopathy and coronary heart disease. The model enables correlation between individual risk factor and the probability of different complications for diabetic patients (Sangi, Win, Shirvani, Namazi-Rad, & Shukla, 2015).

Another study aims to develop multiple neural networks and compare their predictive accuracies with standard statistical model that are being used which is the logistic regression (LR) and quadratic discriminant analysis (QDA) for diagnosis of coronary heart disease (CHD). The assessment reveals that the accuracies of the neural network to be superior (NN = 84.7%, LR = 73.2% and QDA = 58.4%) (Süt & Şenocak, 2007).

Furthermore, a group of researchers developed a new approach to diagnose and classify risk in dengue patients through the implementation of ANN and BIA. The standard method to do so are invasive and time consuming due to the requirement of blood samples. The ANN model that was develop produces and overall prediction accuracy of 96.86% and thus provides an alternative method of diagnosis and classification without the drawback imposed by the standard method (Ibrahim, Faisal, Salim, & Taib, 2010).

CHAPTER 3: METHODOLOGY

3.1 Introduction

This chapter describes the methodology utilized in the production of the classification model for the lipid profile parameters and is illustrated in Figure 3.1.

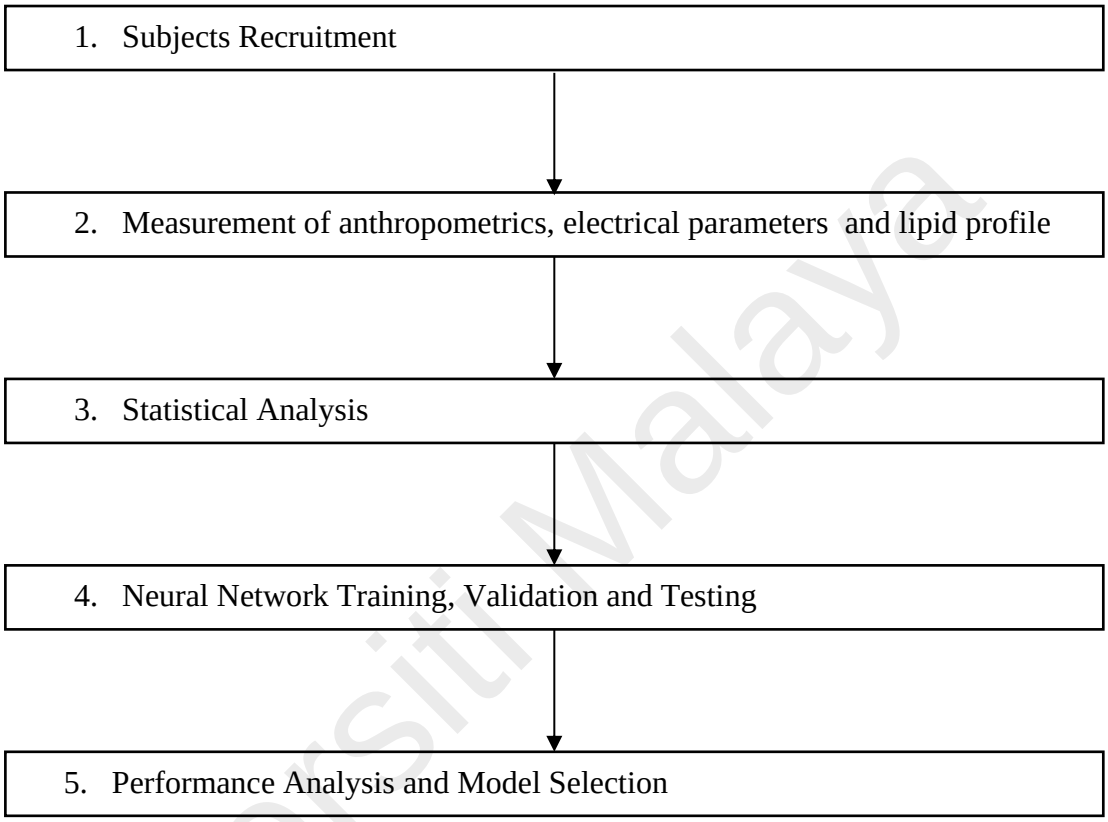

Figure 3.1: Flowchart for the Development of Artificial Neural Network Classification System

3.2 Subjects Recruitment

Subjects are recruited from the Department of Primary Care Medicine in the University Malaya Medical Centre (UMMC) who came for consultation and are required to undergo lipid profile testing. Patient are explained on the objective and protocol of the research through the information sheets provided. Patient consent are taken before initiation of the study.

3.2.1 Subjects Inclusion Criteria

Inclusion criteria includes patients aged of 19 to 80 years old, diagnosed with metabolic or endocrine related diseases such as dyslipidemia, diabetes mellitus, hypertension and any those requiring lipid profile test for further evaluation and diagnosis

3.2.2 Subjects Exclusion Criteria

Exclusion criteria includes chronically ill patients, disabled patients, pregnant or lactating women, and those using medical devices such as ventilators and pacemakers. In lieu of the current COVID-19 circumstances, patients with obvious respiratory symptoms are excluded as well.

3.3 Measurements of Variables

3.3.1 Anthropometric Measurements

Anthropometric measurements that are included are the measurement of subject's height, weight and the calculation of their body mass index (BMI). Subject's height was measured to the nearest 0.1 centimetres (cm) by using a standard wall mounted measuring tape and then converted into metres (m) for the purpose of BMI calculation. Subject's weight was measured to the nearest 0.1 kg by using a Seca Digital Flat Scale.

BMI was calculated through the following equation:

$BMI = kg/m^2$……Equation 3.1

3.3.2 SMARTMF Electrical Parameters

A Bioelectrical Impedance Analysis (BIA) measurement was performed utilizing the SmartMF bioelectrical impedance analyser. The subject was instructed to lie down in a supine position for the placement of the adhesive electrode. The adhesive was then placed on the subject's right hand, one at the metacarpophalangeal joint and another at the wrist.

In addition, two electrodes were then placed on the right foot, between the metatarsophalangeal of the great toe and the second toe and another at ankle joint

Subsequently, the SMARTMF device is switched on and the SMARTMF application on the mobile device is launched. The Bluetooth setting on the smartphone is enabled and the information such as name, height, weight, hours of fasting and gender are being registered through the SMARTMF app. The SMARTMF app is then connected and paired with the SMARTMF device and the electrical parameters and body composition results are displayed within 10 seconds.

3.3.3 Lipid Profile Measurements

The subject was then assisted for the venepuncture procedure that was conducted by the nurses in the Pusat Setempat Pengambilan Darah (PSPD) which serves as a centre for taking blood samples for the outpatient department. Around 10ml of serum is retrieved from the patients and sent to the diagnostic laboratory. Enzymatic analysis is performed on the sample to estimate the lipid levels in the serum samples. The results of the lipid profile are then retrieved from the database with the help of the primary care consultant. A cut off value was set for classification of normal level (0) and abnormal level (1) of the TC, TG, HDL-C and LDL-C based on guideline provided by MIMS for the diagnosis of dyslipidaemia (MIMS, 2020). TC level of equal to or less than 5.2 mmol/L are considered normal and level of more than 5.2 mmol/L are set as abnormal. TG level of equal to or less than 1.7 mmol/L are considered normal and level of more than 1.7 are set as abnormal. HDL-C level of equal to or more than 1.6 mmol/L are considered normal and level of less than 1.6 mmol/L are set as abnormal. LDL-C level of equal to or less than 2.6 mmol/L are considered normal and level of more than 2.6 are set as abnormal. The number 0 for normal and 1 for abnormal represents this status in the ANN.

3.4 Subjects Grouping

87 subjects were recruited and grouped according to their level (normal/abnormal) in all 4 type of parameters in the lipid profile. In each parameter, ANN models are developed in which patients are randomly grouped into training group and testing group consisting 50% of the patient's population respectively.

3.5 Statistical Data Analysis

Statistical analysis utilizing the IBM SPSS Statistics (2020) software (Armonk, NY: IBM) was used to understand the relationship between the variables and to select the predictors to be used in the development of the ANN model. The level of individual lipid profile parameters is set as the dependent variable and measured variable from the BIA and anthropometry are set as independent variables.

A standard descriptive analysis and an independent T-Test was employed to summarize and determine the differences between the mean of the independent variables in the normal group and abnormal group of each lipid profile parameters. Then a simple logistic regression was conducted to select the most appropriate variables from the electrical parameter for the ANN model development. P-value of less than 0.25 is chosen as the significant cut-off point for variables selection as recommended by Hosmer and Lemeshow for screening criterion. This is due the fact that the traditional levels such as p-value of less than 0.05 has been proven often to fail in identifying important variable for building a model (Hosmer Jr, Lemeshow, & Sturdivant, 2013).

3.6 Development of Artificial Neural Network

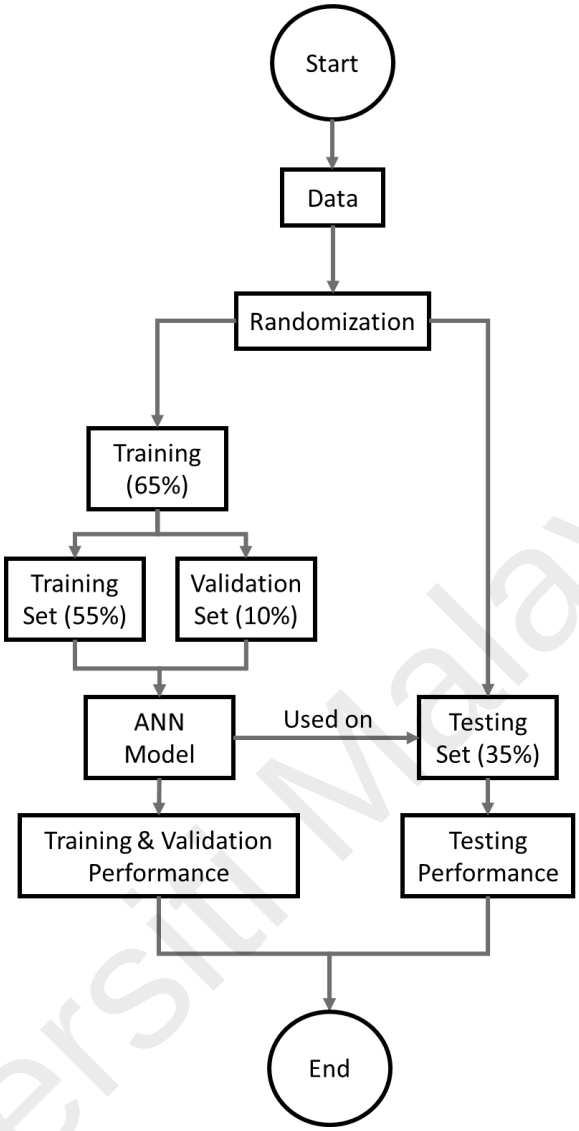

Figure 3.2: Flowchart ANN Development

Description of the network development and characteristics to be used to classify the lipid profile levels is described in this section and Figure 3.2 shows the flowchart of the development. A supervised learning method employing a multi-layered feed forward perceptron with backpropagation was utilized.

Learning algorithm for the classification that was selected includes the Levenberg-Marquardt, Resilient and Scaled Conjugate Gradient. ANN models are produced for each algorithm and each parameters of the lipid profile. As sample size was small, each model

independently uses randomized data for training and testing from the 87 data sets that was available.

The Neural Pattern Recognition Toolbox (nprtool) by MATLAB. (2019). *version 9.7.0.11 (R2019b)* (Natick, Massachusetts: The MathWorks Inc) was used for the neural network development. The methods used in the tools are explained in the subsequent section.

3.6.1 Data Division

Datasets are randomly divided into three subsets. The first subset of data is the training set used to compute and update the weight and biases of the network. This subset is presented to the network during the training stage and the network will be adjusted according to the error produced. The second subset is the validation set. This subset is utilized to measure the generalization of the network and halt training when improvement of generalization stops. This happens when the network starts to overfit the data and validation error rises. The final subset of data is the test set, that has no impact on the training process and thus provides an independent performance measurement after training. 55% of the total datasets is used for the training set, 10% of the total datasets is used for the validation set and the remaining 35% of the total datasets is used for the training set.

3.6.2 Performance Function

The training process of the neural network involves tuning of the values of weight and biases of the network to optimize its performance and this is defined by the network performance function. For the purpose of developing a neural network model for classification the cross-entropy function is utilized. The function returns a result that heavily penalizes outputs that are extremely inaccurate (y near 1-t), with very little

penalty for fairly correct classifications (y near t). Minimizing cross-entropy leads to good classifiers.

3.6.3 Neural Network Structure

3.6.3.1 Total Cholesterol

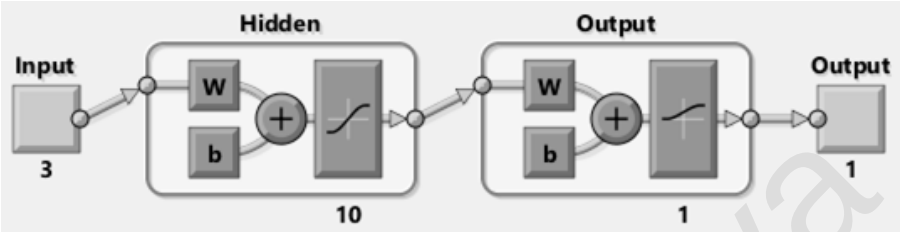

Figure 3.3: TC Model Network Structure

The neural network structure for the development of TC models is shown in Figure 3.3. It consists of 3 input layers and 1 output layer. The 3 input are the SMARTMF electrical measurements of body capacitance, resistance and reactance. The output is the TC level class.

3.6.3.2 Triglycerides

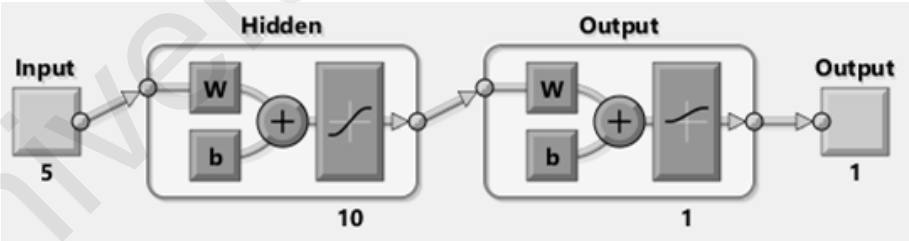

Figure 3.4: TG Model Network Structure

The neural network structure for the development of TG neural network is shown in Figure 3.4. It consists of 4 input layers and 1 output layer. The 5 input are the SMARTMF electrical measurements of impedance at 50 ,100 and 200 kHz, resistance and reactance. The output is the TG level class.

3.6.3.3 High Density Lipoprotein Cholesterol

Figure 3.5: HDL Model Network Structure

The neural network structure for the development of HDL-C neural network is shown in Figure 3.5. It consists of 4 input layers and 1 output layer. The 4 input are the SMARTMF electrical measurements of impedance at 5, 50 ,100 and 200 kHz. The output is the HDL-C level class.

3.6.4 Training of Network

3.6.4.1 Supervised Multi-layered Perceptron

The system of a multilayer feedforward neural network (MFNN) basic building block consists of the three layers; one input layer, one hidden layer and one output layer. The neuron or processing unit consist of one or multiple inputs formed by the vector $\bar{X} = (X_1, X_2 \ldots X_i)$ where i is the number of input variables selected through the statistical analysis.

The output of the networks of the lipid profile (LP = TC, TG, HDL-C and LDL-C) is defined by the following equation (Galushkin, 2007):

Output layer,

$LP = f_2(A_{k2})$…… Equation 3.2

Hidden Layer,

$H = f_1(A_{k1})$……Equation 3.3

Activation equation,

$A_k = \sum_i X_i W_{ki} + \theta_i$ …… Equation 3.4

Where;

X = input neuron

W_{ki}= weight or connection strength

f_1/f_2= transfer function of respective layer

In the hidden layer, a tangent sigmoid activation function were used and in the output layer a logistic sigmoid activation function were used with the following equation (Fukunaga, 1990):

Tangent sigmoid = $f_1 = \frac{2}{(1+e^{((-2X)-1)})}$…… Equation 3.5

Logistic sigmoid = $f_2 = \frac{1}{1+e^{-X}}$…… Equation 3.6

3.6.4.2 Backpropagation Algorithm

Each iteration will undergo through a backpropagation algorithm to adjust the weight for each neuron accordingly based on the differences of the target and output of the model. 3 different backpropagations algorithm are employed, the scaled conjugate gradient (SCG), Levenberg-Marquardt (LM) and Resilient (RB). The initialization of weight and bias for the purpose of training is automatically set by the toolbox.

3.6.5 Network Generalization

Generalization of any neural network is sensitive to the number of hidden neurons in the hidden layers. A low number of neurons would lead to and underfit network and a high number of neurons could lead to overfitting. Neural networks are usually optimized

through training with different number of neurons. In this study the number of neurons is set at 10 (default) and optimization is not done in the context of number of neurons as the aim of the research is to only show the ability of ANN for the classification of the lipid profile measurements.

Generalization improvements in the toolbox uses the 'early stopping' method by default. The error on the validation set is monitored through the training process. When validation error indicated by the cross-entropy value increase for a set number of iterations, the training process will be halted. In this study, the number of iterations set for the method is set at 6 (default). The maximum number of epoch is set at 1000 (default). The actual number of epoch used for the network model are dependent on the 'early stopping' method.

3.6.6 Performance Analysis and Model Selection

The performance of the neural networks models based on the individual lipid profile parameters were evaluated by its accuracy, sensitivity and specificity on the testing subset of the data. There are four important test outcomes; true positive (TP), true negative (TN), false-positive (FP) and false negative (FN). The accuracy of the model represents the proportion of true positive results in the selected sample of population, Accuracy = (TN+TP)/(TN+TP+FN+FP). The sensitivity of the model represents the proportion of correctly identified positive results in the number of all actual positive results, Sensitivity = TP/(TP+FN), and the specificity of the model represent the proportion of correctly identified negative results in the number of all actual negative results, Specificity = TN/(TN+FP). A gold standard test or diagnostic test in the clinical field are often those with high specificity and sensitivity. It is not uncommon for test to have high sensitivity or specificity but not both especially in screening tests. The trade-offs from this drawback is that screening tests are inexpensive, quick and often non-invasive or minimally invasive

as compared to diagnostic tests (Trevethan, 2017). However, the impact of having a low sensitivity or specificity in a screening test must be thoroughly examined due to possible serious consequences for both individuals and the healthcare system (Mallett, Halligan, Thompson, Collins, & Altman, 2012; Pewsner et al., 2004).

CHAPTER 4: RESULTS AND DISCUSSION

4.1 Introduction

This chapter presents the results of the statistical analysis and the development process of the artificial neural network. Section 4.2 demonstrates the demographic distributions of the collected data in the study. A T-Test analysis for each lipid parameters are discussed in section 4.3. Furthermore, a logistic regression for each independent variable is analysed and deliberated in section 4.4 for the purpose of variable selection for the ANN development. In section 4.5, the performance of developed ANN of different algorithms is compared.

4.2 Demographic Data

A total of 87 patients has been recruited as subjects for the purpose of the study. All patients that were recruited are used for the purpose of the analysis, hence 87 sets of data are available. From the 87 data sets, 35 (40%) are males and 52 (60%) are females, aged between 19 to 80 years old (mean 58.43, SD ± 1.39).

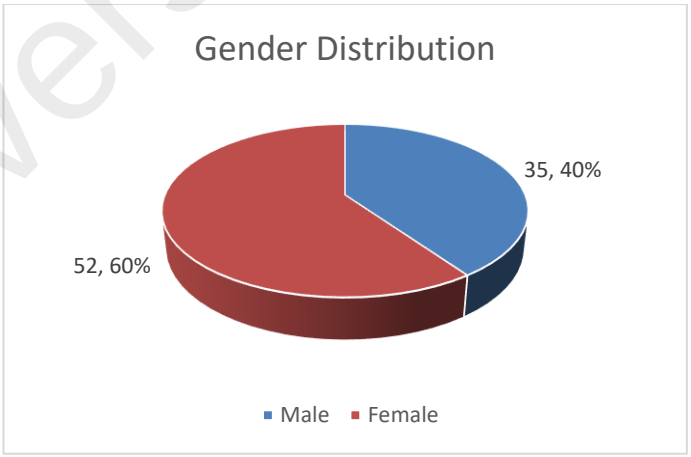

Figure 4.1: Gender Distribution in Data Sets

4.3 T-Test Analysis

The patients are grouped according to their lipid levels (Normal or Abnormal) for each measurement in the lipid profile (Table 4.1) based on MIMS guidelines on the diagnosis of dyslipidaemia (MIMS, 2020). A descriptive analysis and T-Test analysis of the anthropometric and SmartMF bioelectrical parameters is shown in Table 4.2 to Table 4.4. As a screening criterion, alpha is set at 0.25 thus a p-value of <0.25 is considered significant.

Table 4.1: Grouping Criteria for Lipid Profile

Lipid Parameters	Normal Level (mmol/L)	Abnormal Level
TC	≤ 5.2	> 5.2
TG	≤ 1.7	> 1.7
HDL-C	≥ 1.6	<1.6
LDL-C	≤ 2.6	>2.6

Table 4.2: Parameters based on TC Group

Total Cholesterol			
Parameter (units)	Normal Group (Mean ± SD)	Abnormal Group (Mean ± SD)	P-Value
	n = 52 (M=25, F = 27)	n = 35 (M=10,F=25)	
Age (years)	59.48 ± 11.92	57.03 ± 14.15	0.386
Height (cm)	162.96 ± 9.87	158.09 ± 6.26	0.011*
Weight (kg)	68.19 ± 13.21	67.09 ± 13.15	0.702
Phase Angle (°)	6.32 ± 1.55	6.68 ± 1.33	0.254
Body Capacitance (pF)	56.84 ± 17.62	50.96 ± 11.63	0.087*
Resistance (ohms)	482.697 ± 192.72	406.95 ± 237.37	0.105*
Reactance (ohms)	129.63 ± 167.17	223.49 ± 249.34	0.038
Impedance at 5kHz (ohms)	573.32 ± 77.36	584.23 ± 93.08	0.554
Impedance at 50 kHz (ohms)	554.878 ± 76.56	568.56 ± 87.50	0.443
Impedance at 100 kHz (ohms)	525.42 ± 75.75	537.45 ± 87.36	0.497
Impedance at 200 kHz (ohms)	494.02 ± 73.31	504.09 ± 86.25	0.561

*p-value < 0.25

Table 4.3: Parameters based on TG Group

Triglycerides			
Parameter (units)	Normal Group (Mean ± SD)	Abnormal Group (Mean ± SD)	P-Value
	n = 62 (M=26, F = 36)	n = 25 (M=9, F=16)	
Age (years)	58.53 ± 12.51	58.4 ± 13.90	0.966
Height (cm)	161.87 ± 9.38	158.84 ± 7.28	0.151*
Weight (kg)	66.34 ± 13.60	71.24 ± 11.37	0.115*
Phase Angle (°)	6.37 ± 1.52	6.68 ± 1.33	0.366
Body Capacitance (pF)	54.98 ± 17.19	53.23 ± 11.35	0.641
Resistance (ohms)	486.28 ± 198.45	367.78 ± 230.70	0.018
Reactance (ohms)	141.46 ± 187.28	231.69 ± 244.86	0.067*
Impedance at 5kHz (ohms)	583.92 ± 88.76	562.31 ± 68.69	0.278
Impedance at 50 kHz (ohms)	568.29 ±83.05	540.77 ± 73.34	0.152*
Impedance at 100 kHz (ohms)	538.45 ± 81.44	509.94 ± 75.29	0.134*
Impedance at 200 kHz (ohms)	506.71 ± 78.28	476.61 ± 76.17	0.106*

*p-value <0.25

Table 4.4: Parameters based on HDL-C Group

High Density Lipoprotein Cholesterol			
Parameter (units)	Normal Group (Mean ± SD)	Abnormal Group (Mean ± SD)	P-Value
	n = 23 (M=6, F = 17)	n = 64 (M= 29, F = 35)	
Age (years)	59.61 ± 10.70	58.09 ± 13.59	0.630
Height (cm)	157.04 ± 4.94	162.42 ± 9.57	0.012*
Weight (kg)	60.39 ± 8.82	70.39 ± 13.45	0.001*
Phase Angle (°)	6.41 ±1.66	6.48 ± 1.40	0.830
Body Capacitance (pF)	52.35 ± 17.89	55.24 ± 14.89	0.453
Resistance (ohms)	496.35 ± 202.87	436.37 ± 216.89	0.251
Reactance (ohms)	165.32 ± 234.54	168.14 ± 199.72	0.956
Impedance at 5kHz (ohms)	617.60 ± 80.58	563.37 ± 80.64	0.007*
Impedance at 50 kHz (ohms)	598.67 ± 87.90	546.66 ± 74.29	0.008*
Impedance at 100 kHz (ohms)	567.19 ± 85.78	516.98 ± 74.60	0.009*
Impedance at 200 kHz (ohms)	532.62 ± 81.81	485.65 ± 73.95	0.013*

*p-value < 0.25

Table 4.5: Parameters based on LDL-C Group

Low Density Lipoprotein Cholesterol			
Parameter (units)	Normal Group (Mean ± SD)	Abnormal Group (Mean ± SD)	P-Value
	n = 33 (M=16, F = 17)	n = 54 (M=19,F=35)	
Age (years)	61.42 ± 10.50	56.70 ± 13.87	0.960
Height (cm)	161.61 ± 9.73	160.63 ± 8.41	0.622
Weight (kg)	66.06 ± 10.46	68.78 ± 14.51	0.352
Phase Angle (°)	6.39 ± 1.42	6.51 ± 1.51	0.715
Body Capacitance (pF)	54.67 ± 13.44	54.36 ± 17.03	0.929
Resistance (ohms)	462.85 ± 220.47	445.73 ± 211.40	0.719
Reactance (ohms)	154.61 ± 185.12	175.20 ± 222.26	0.657
Impedance at 5kHz (ohms)	581.09 ± 75.48	575.64 ± 88.96	0.770
Impedance at 50 kHz (ohms)	558.97 ± 79.02	561.24 ± 82.80	0.900
Impedance at 100 kHz (ohms)	529.67 ± 78.04	530.61 ± 82.44	0.958
Impedance at 200 kHz (ohms)	498.32 ± 75.10	497.91 ± 80.94	0.981

*p-value < 0.25

Results for the independent t-test analysis for the TC parameter was tabulated and shown in Figure 4.2. The analysis shows significant difference between the TC groups in terms of the height, body capacitance and resistance parameters. All three parameters are

negatively correlated with the TC levels. A study conducted on the relationship between TC and body capacitance has shown a positive correlation (Mohktar, Ibrahim, & Ismail, 2007) thus contradicts with the outcome of the analysis.

Results for the independent t-test analysis for the TG parameter was tabulated and shown in Table 4.3. The analysis shows significant difference between the TG groups in terms of the height, weight, reactance, impedance at 50 kHz, impedance at 100 kHz and impedance at 200 kHz parameters. The height and all three impedance parameters are negatively correlated with the TG levels, On the other hand, the weight and reactance parameters are positively correlated with the TG levels.

Results for the independent t-test analysis for the HDL-C parameter was tabulated and shown in Table 4.4. The analysis shows significant difference between the HDL-C groups in terms of the height, weight, and all 4 of the impedance parameters. The height and weight are positively correlated with the HDL-C level. Conversely, all 4 impedance parameters are negatively correlated with the HDL-C level.

Results for the independent t-test analysis for the LDL-C parameter was tabulated and shown in Table 4.5. The analysis indicates that no parameters show significant difference between the LDL-C group.

4.4 Logistic Regression Analysis and Variable Selection

The result from the simple logistic regression analysis is shown in Table 4.6 to Table 4.9. Only the bioelectrical parameters are analysed. Two important results are shown in each analysis which is the p-value determining the significance of the independent variables or predictors and the odds ratio (OR) of individual predictors. As a screening criterion, alpha is set at 0.25 thus a p-value of <0.25 is considered significant.

Table 4.6: Logistic Regression Result for TC

Total Cholesterol		
Variable	**P-Value**	**OR**
Phase Angle (°)	0.252	1.191
Body Capacitance (pF)	0.093*	.973
Impedance at 5 kHz (ohms)	0.549	1.002
Impedance at 50 kHz (ohms)	0.439	1.002
Impedance at 100 kHz (ohms)	0.492	1.002
Impedance at 200 kHz (ohms)	0.556	1.002
Resistance (ohms)	0.108*	.998
Reactance (ohms)	0.044*	1.002

*p-value < 0.25

Table 4.7: Logistic Regression Result for TG

Triglycerides		
Variable	**P-Value**	**OR**
Phase Angle (°)	0.363	1.162
Body Capacitance (pF)	0.637	0.993
Impedance at 5 kHz (ohms)	0.277	0.997
Impedance at 50 kHz (ohms)	0.154*	0.995
Impedance at 100 kHz (ohms)	0.137*	0.995
Impedance at 200 kHz (ohms)	0.109*	0.995
Resistance (ohms)	0.022*	0.998
Reactance (ohms)	0.073*	1.002

*p-value < 0.25

Table 4.8: Logistic Regression Result for HDL-C

High Density Lipoprotein Cholesterol		
Variable	**P-Value**	**OR**
Phase Angle (°)	0.828	1.037
Body Capacitance (pF)	0.449	1.013
Impedance at 5 kHz (ohms)	0.010*	0.992
Impedance at 50 kHz (ohms)	0.011*	0.992
Impedance at 100 kHz (ohms)	0.013*	0.992
Impedance at 200 kHz (ohms)	0.017*	0.992
Resistance (ohms)	0.252	0.999
Reactance (ohms)	0.955	1.000

*p-value < 0.25

Table 4.9: Logistic Regression Results for LDL-C

Low Density Lipoprotein Cholesterol		
Variable	**P-Value**	**OR**
Phase Angle (°)	0.711	1.058
Capacitance (pF)	0.928	0.999
Impedance at 5 kHz (ohms)	0.767	1.000
Impedance at 50 kHz (ohms)	0.898	1.000
Impedance at 100 kHz (ohms)	0.957	1.000
Impedance at 200 kHz (ohms)	0.981	1.000
Resistance (ohms)	0.716	1.000
Reactance (ohms)	0.653	1.000

*p-value < 0.25

The logistic regression analysis result indicates several significant predictors that can be utilized for the purpose of the development of ANN. For ANN model of TC, body capacitance, resistance and reactance parameters will be utilized as inputs. For ANN model of TG, impedance at 50,100 and 200 kHz, resistance and reactance parameters will be utilized as inputs. For ANN model of HDL-C, impedance at 5, 50, 100 and 200 kHz parameters will be utilized as inputs. For LDL-C, there are no significant predictor determined from the analysis. Hence, an ANN model could not be developed for the LDL-C parameter.

4.5 Development of Artificial Neural Network and Model Selection

Artificial Neural Network employing the MFNN technique was developed for the TC, TG and HDL-C parameters using the scaled conjugate gradient (SCG), Levenberg Marquardt (LM) and Resilient (RB) algorithm. 10 neural networks were produced using each algorithm and an average of the performance were calculated. Sensitivity indicates the percentage of true positive (abnormal level) from all actual positive result, specificity indicates the percentage of true negative (normal level) from all actual negative result and accuracy indicates the percentage of total true negative and true positive from all result.

4.5.1 Total Cholesterol Models

4.5.1.1 TC – Scaled conjugate Gradient (SCG)

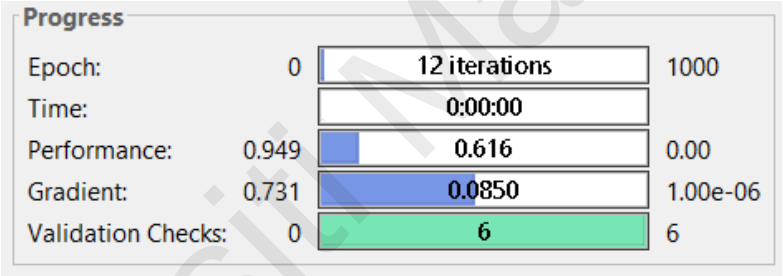

Figure 4.2: Training Progress Summary (TC-SCG)

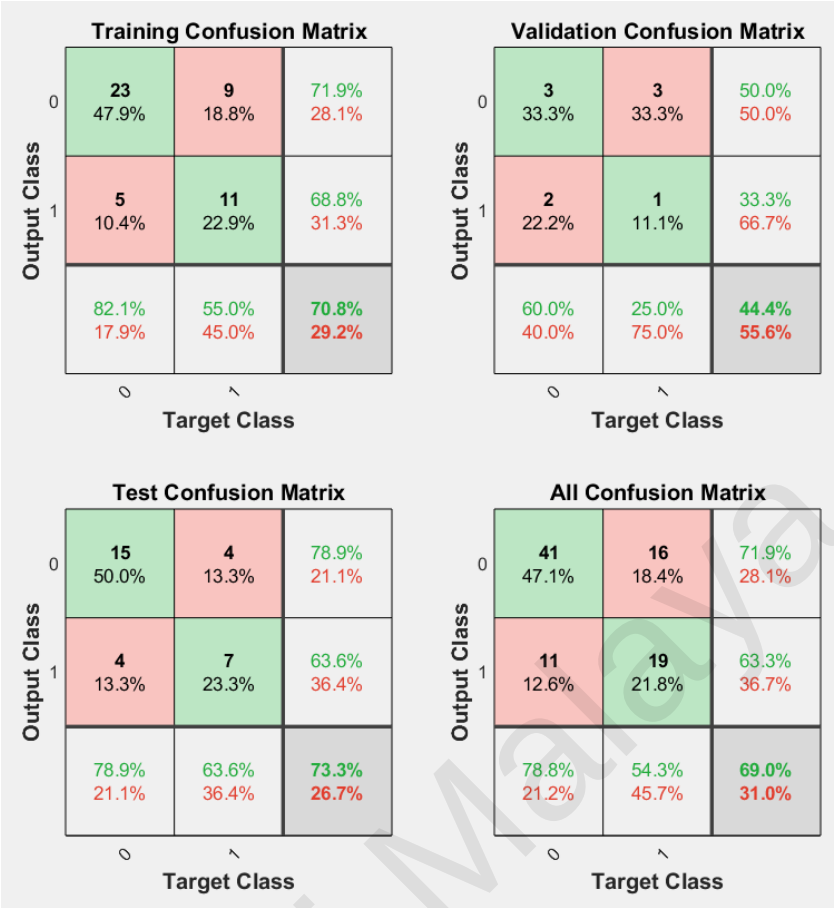

Figure 4.3: Confusion Plot (TC-SCG)

The training process for the TC model for SCG stops at epoch 12 with a cross entropy value of 0.616 and a gradient of 0.0850 (Figure 4.2). Based on the confusion plot in Figure 4.3, training of the network produces an accuracy of 70.8%, sensitivity of 55.5% and specificity of 82.1%. Performance of testing shows an accuracy of 73.3%, sensitivity of 63.6% and specificity of 78.9%.

4.5.1.2 TC – Resilient Backpropagation

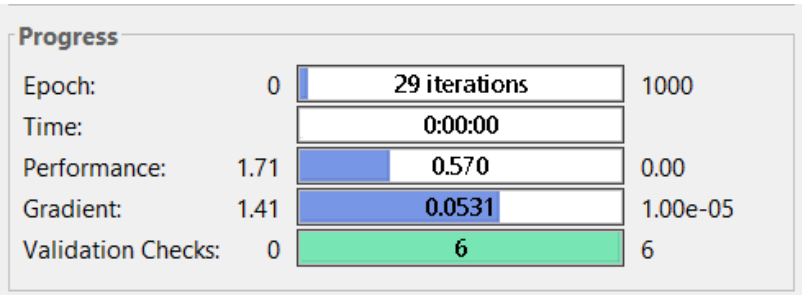

Figure 4.4: Training Progress Summary (TC-RB)

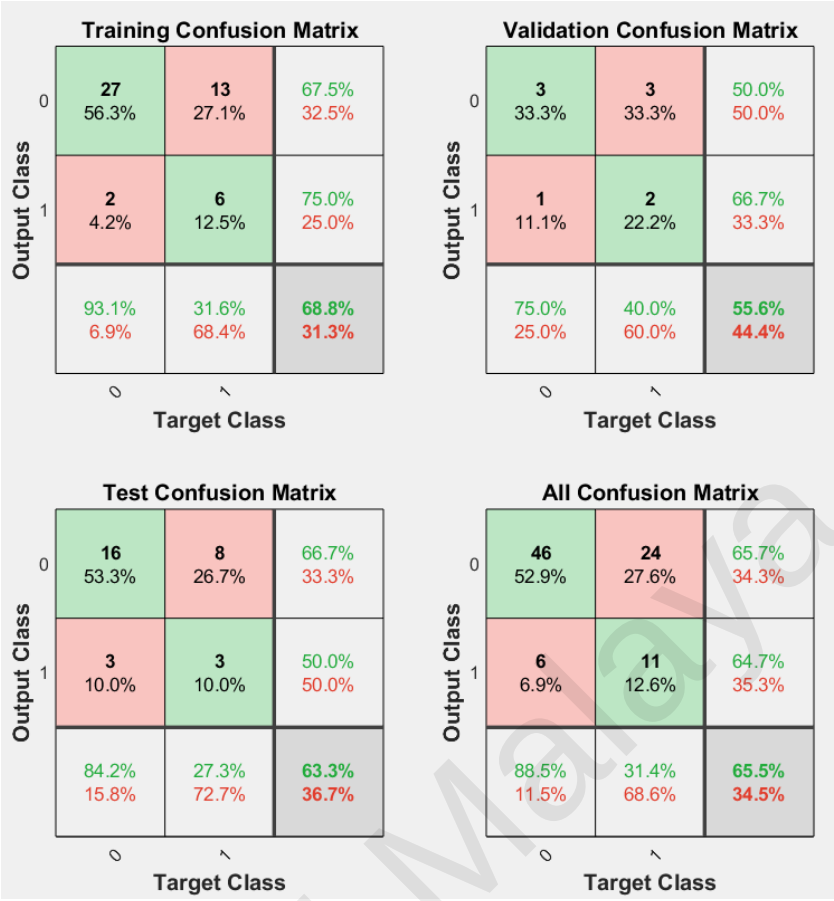

Figure 4.5: Confusion Plot (TC-RB)

The training process for the TC model for RB stops at epoch 29 with a cross entropy value of 0.570 and a gradient of 0.0531 (Figure 4.4). Based on the confusion plot in Figure 4.5, training of the network produces an accuracy of 68.8%, sensitivity of 31.6% and specificity of 93.1%. Performance of testing shows an accuracy of 63.3%, sensitivity of 27.3% and specificity of 84.2%.

4.5.1.3 TC – Levenberg-Marquardt (LM)

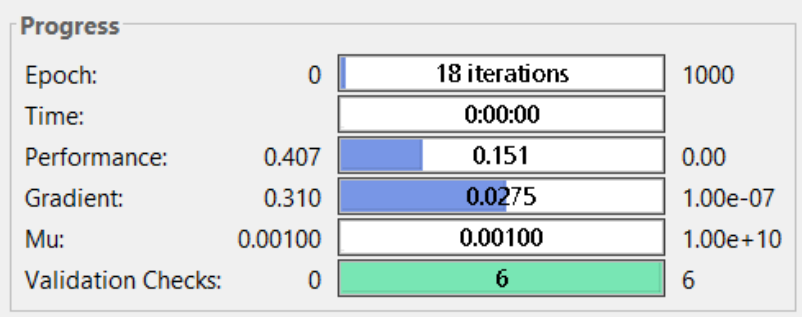

Figure 4.6: Training Progress Summary (TC-LM)

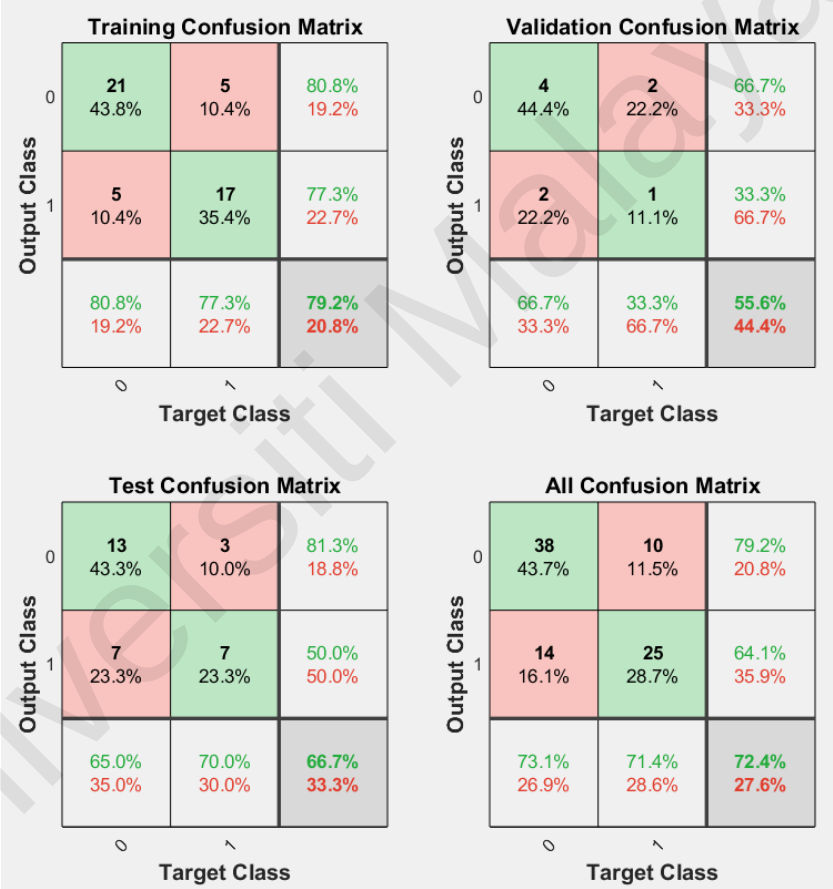

Figure 4.7: Confusion Plot (TC-LM)

The training process for the TC model for LM stops at epoch 18 with a cross entropy value of 0.151 and a gradient of 0.0275 (Figure 4.6). Based on the confusion plot in Figure 4.7, training of the network produces an accuracy of 79.2%, sensitivity of 77.3% and specificity of 80.8%. Performance of testing shows an accuracy of 66.7%, sensitivity of 65.0% and specificity of 70.0%.

4.5.2 Triglycerides Models

4.5.2.1 TG – Scaled Conjugate Gradient

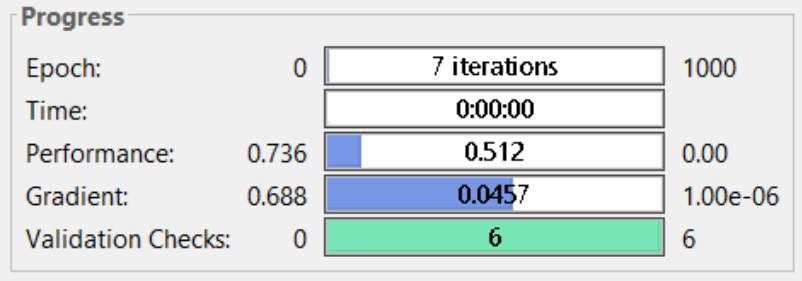

Figure 4.8: Training Progress Summary (TG-SCG)

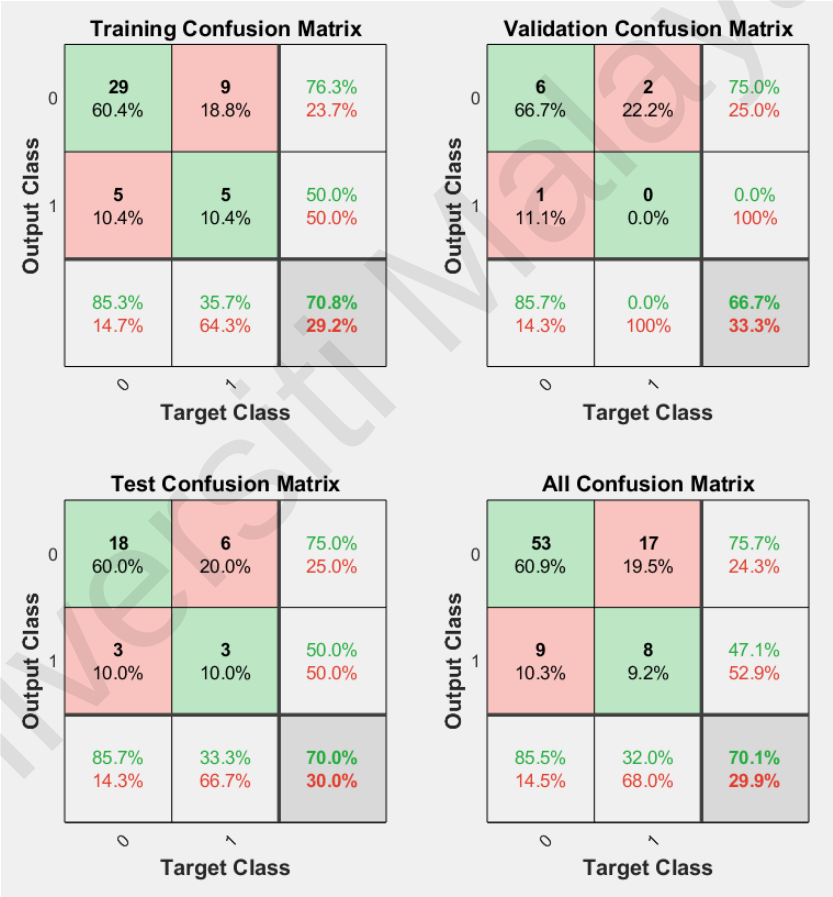

Figure 4.9: Confusion Plot (TG-SCG)

The training process for the TG model for SCG stops at epoch 7 with a cross entropy value of 0.512 and a gradient of 0.0457 (Figure 4.8). Based on the confusion plot in Figure 4.9, training of the network produces an accuracy of 70.8%, sensitivity of 35.7% and

specificity of 85.3%. Performance of testing shows an accuracy of 70%, sensitivity of 33.3% and specificity of 86.7%

4.5.2.2 TG – Resilient Backpropagation (RB)

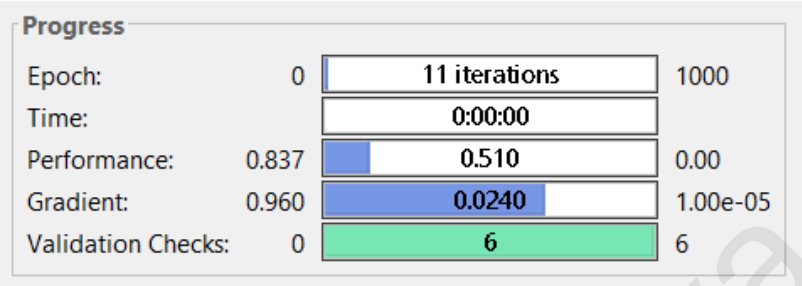

Figure 4.10: Training Progress Summary (TG-RB)

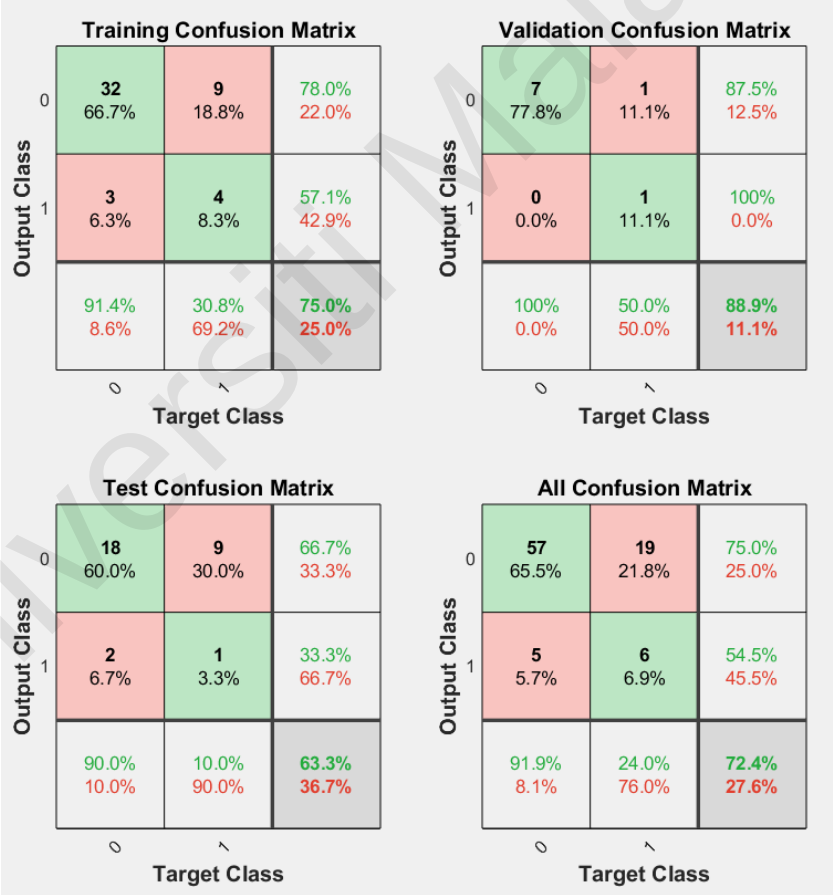

Figure 4.11: Confusion Plot (TG-RB)

The training process for the TG model for RB stops at epoch 11 with a cross entropy value of 0.510 and a gradient of 0.0240 (Figure 4.10). Based on the confusion plot in Figure 4.11, training of the network produces an accuracy of 75.0%, sensitivity of 30.8%

and specificity of 91.4%. Performance of testing shows an accuracy of 63.3%, sensitivity of 10% and specificity of 90%.

4.5.2.3 TG – Levenberg-Marquardt

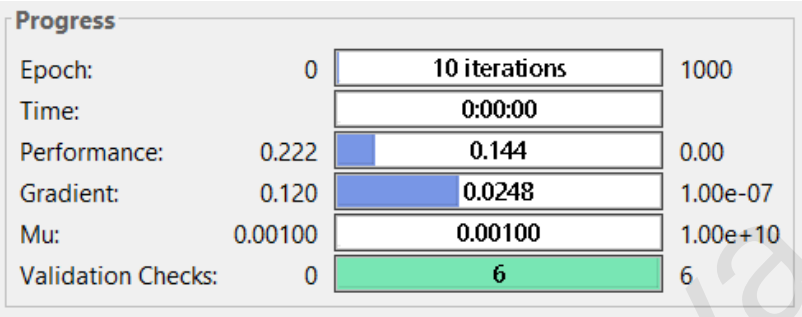

Figure 4.12: Training Progress Summary (TG-LM)

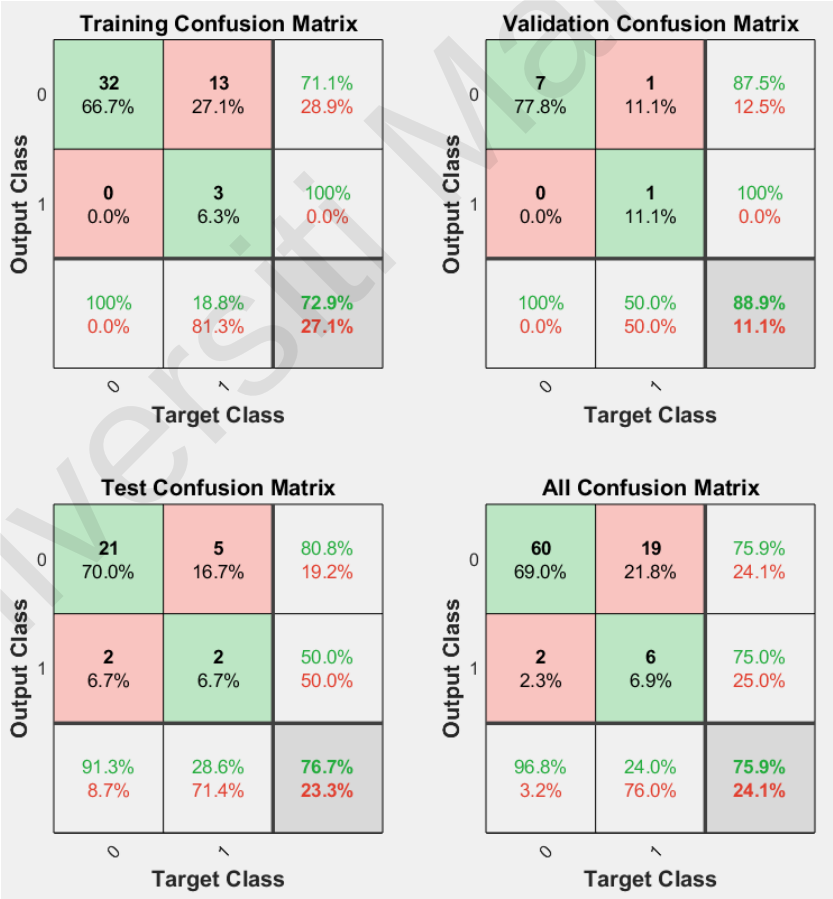

Figure 4.13: Confusion Plot (TG-LM)

The training process for the TG model for LM stops at epoch 10 with a cross entropy value of 0.144 and a gradient of 0.0248 (Figure 4.12). Based on the confusion plot in

Figure 4.13, training of the network produces an accuracy of 72.9% sensitivity of 18.0% and specificity of 100%. Performance of testing shows an accuracy of 76.7%, sensitivity of 28.6 and specificity of 91.3%.

4.5.3 High Density Lipoprotein Cholesterol Models

4.5.3.1 HDL - Scaled Conjugate Gradient (SCG)

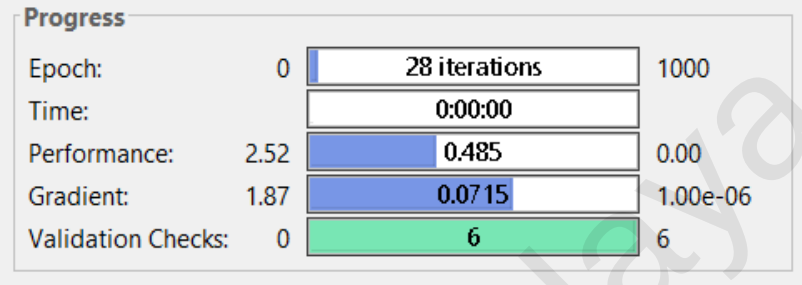

Figure 4.14: Training Progress Summary (HDL-SCG)

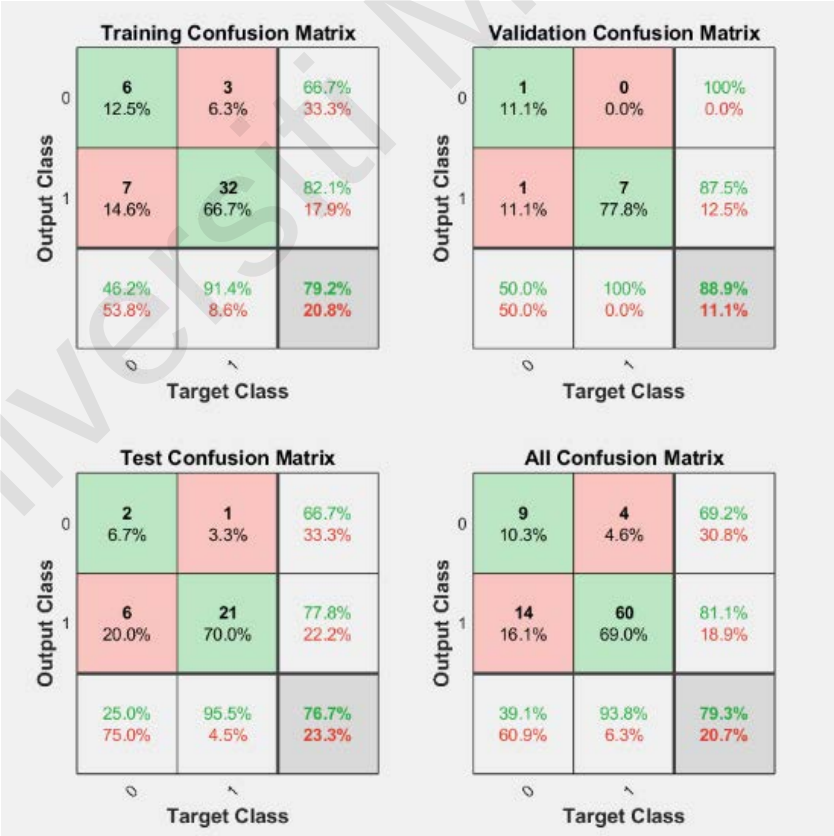

Figure 4.15: Confusion Plot (HDL-SCG)

The training process for the HDL-C model for SCG stops at epoch 28 with a cross entropy value of 0.485 and a gradient of 0.0715 (Figure 4.14). Based on the confusion

plot in Figure 4.15, training of the network produces an accuracy of 79.2%, sensitivity of 91.4% and specificity of 46.2%. Performance of testing shows an accuracy of 76.7%, sensitivity of 95.5% and specificity of 25.0%.

4.5.3.2 HDL - Resilient Backpropagation (RB)

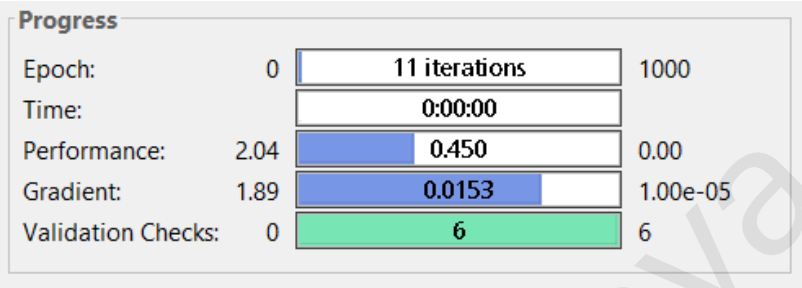

Figure 4.16:Training Progress Summary (HDL-RB)

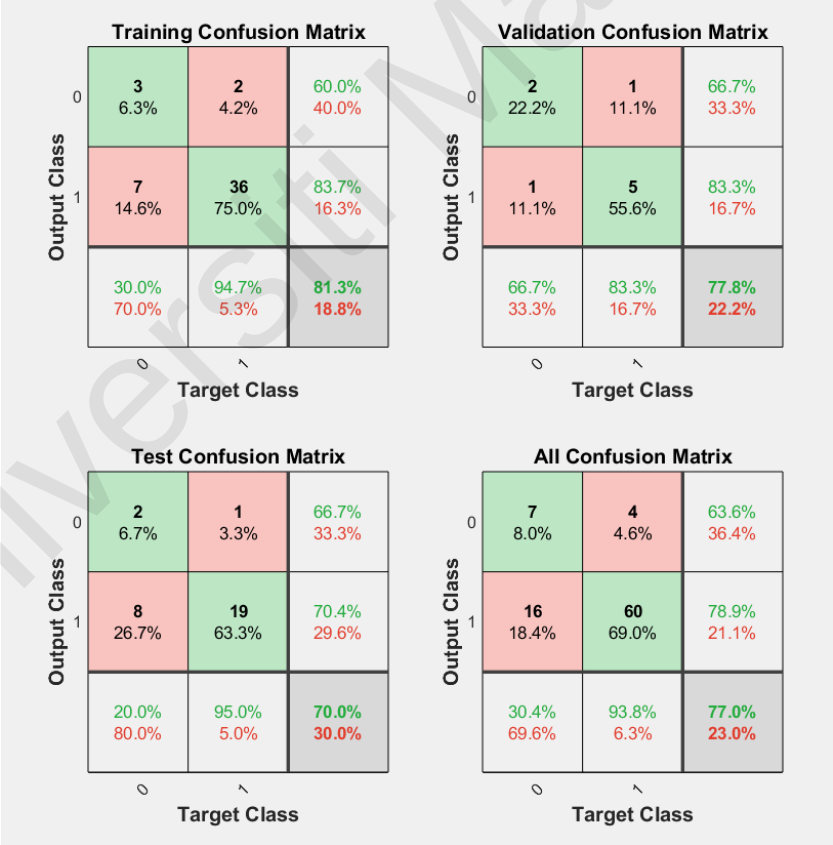

Figure 4.17: Confusion Plot (HDL-RB)

The training process for the HDL-C model for RB stops at epoch 11 with a cross entropy value of 0.450 and a gradient of 0.0153 (Figure 4.16). Based on the confusion

plot in Figure 4.17, training of the network produces an accuracy of 81.3%, sensitivity of 94.7% and specificity of 30.0%. Performance of testing shows an accuracy of 70%, sensitivity of 95% and specificity of 20%.

4.5.3.3 HDL – Levenberg-Marquardt (LM)

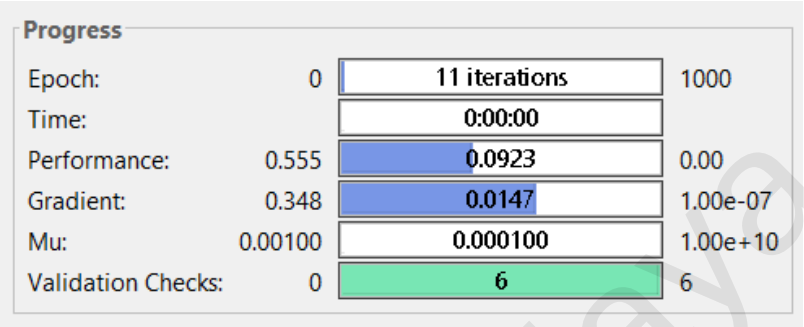

Figure 4.18: Training Progress Summary (HDL-LM)

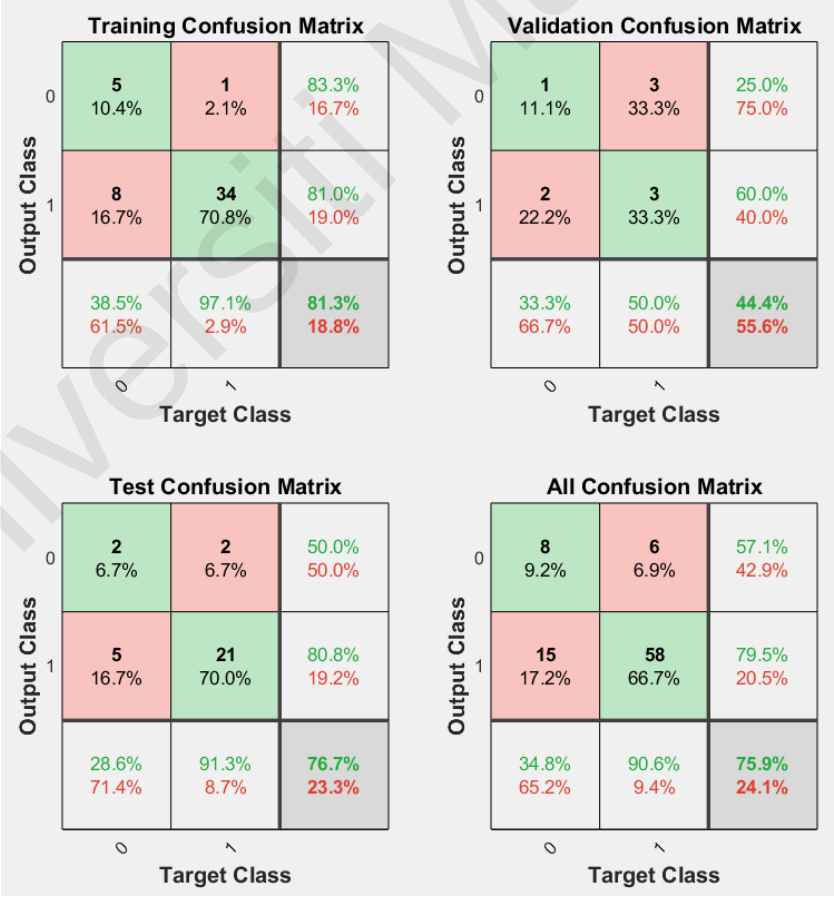

Figure 4.19: Confusion Plot (HDL-LM)

The training process for the HDL-C model for LM stops at epoch 11 with a cross entropy value of 0.0923 and a gradient of 0.0147. Based on the confusion plot in Figure 4.19, training of the network produces an accuracy of 81.3%, sensitivity of 97.1% and specificity of 38.5%. Performance of testing shows an accuracy of 76.6%, sensitivity of 91.3% and specificity of 28.6%.

4.5.4 Models Testing Performance Comparison

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)
TC-SCG	73.3	63.6	78.9
TC-RB	63.3	27.3	84.2
TC-LM	66.7	65.0	70.0
TG-SCG	70.0	33.3	86.7
TG-RB	63.3	10.0	90.0
TG-LM	76.7	28.6	91.3
HDL-SCG	76.7	95.5	25.0
HDL-RB	70.0	95.0	20.0
HDL-LM	76.6	91.3	28.6

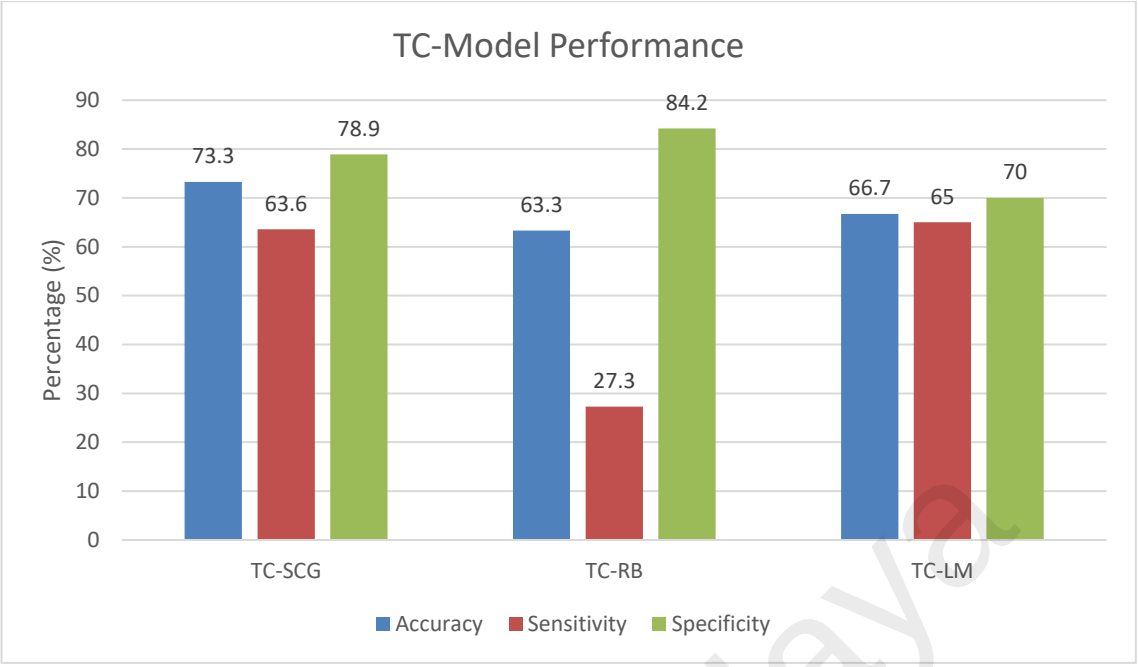

Figure 4.20: TC Models Testing Performance

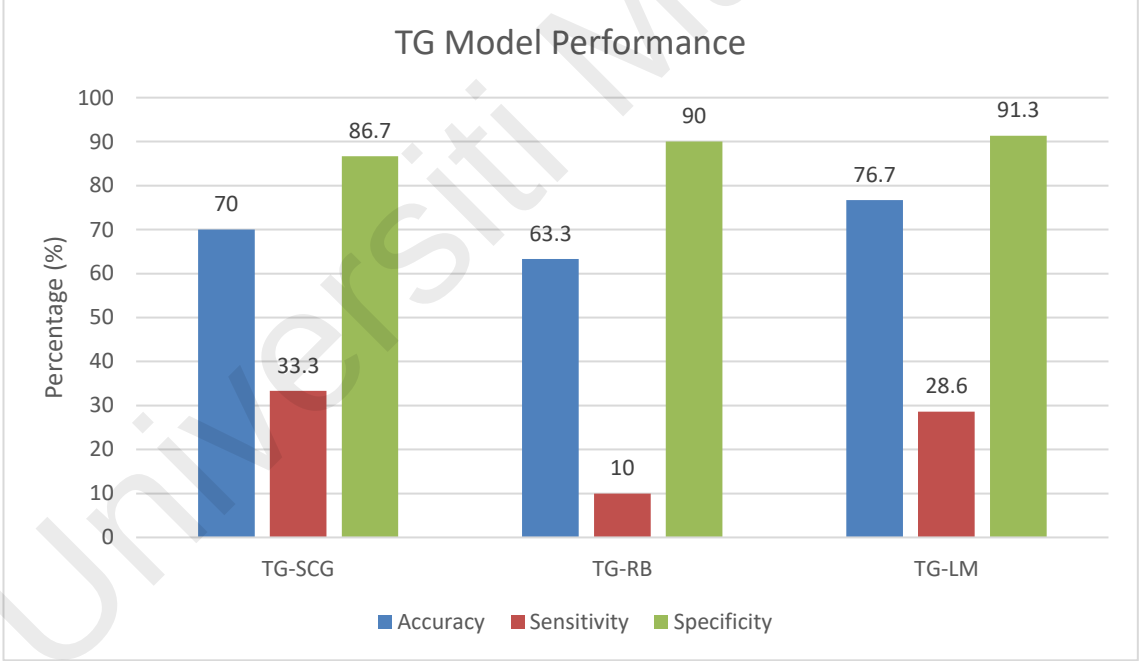

Figure 4.21: TG Models Performance

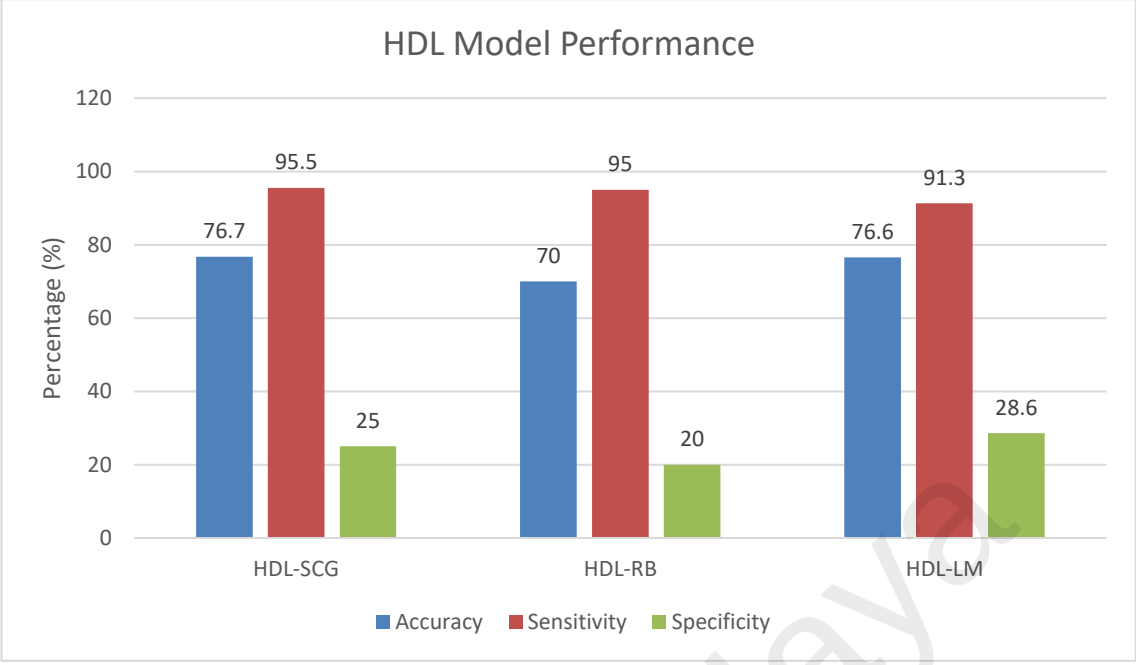

Figure 4.22: HDL Models Performance

Figure 4.20 – 4.22 summarizes the testing performance of the models utilising different backpropagation methods for TC, TG and HDL-C measurements of the lipid profile. The best models for the lipid profile parameters are as following, Total Cholesterol-Scaled Conjugate Gradient (TC-SCG) model (Accuracy, 73.3%; Sensitivity, 63.6%; Specificity, 78.9%), Triglyceride-Levenberg-Marquardt (TG-LM) Model (Accuracy, 76.7%; Sensitivity, 28.6%; Specificity, 91.3%) and High-Density Lipoprotein Cholesterol-Resilient (HDL-SCG) Model (Accuracy, 76.7%; Sensitivity, 95.5%; Specificity, 25.0%). It can be observed that the models for TC and TG have high specificity but low sensitivity. A high specificity test is suitable for ruling in conditions, in this case ruling in abnormal level of TC and TG. On the other hand, the models for HDL have high sensitivity but low specificity. A high sensitivity test is suitable for ruling out conditions, in this case ruling out abnormal level of HDL. Thus, the use of the models can be situational and requires further studies in its suitability of its application in the clinical field.

CHAPTER 5: CONCLUSION AND FUTURE WORK

Dyslipidaemia is a state of abnormal levels of lipids in the blood. It is a risk factor to a multitude of conditions especially in atherosclerotic cardiovascular disease (Kopin & Lowenstein, 2017). The standard test used in determining the levels of the lipid constituents is called a lipid profile test that requires invasive blood sampling for laboratory-based quantification of said levels. The procedure of blood sampling or venepuncture comes with its own risks and complication making screening unappealing (Galena, 1992). This is reflected by the data in Malaysia that shows only 19.2% of those with abnormal cholesterol level are aware that they are affected by the condition (Mat Rifin et al., 2018). Thus, non-invasive techniques to screen dyslipidaemia are in need of research.

BIA is a non-invasive technique that measure bioelectrical parameters from the body then deriving the body composition parameters such as fat mass. This technique has been studied in its use for healthcare screening purposes. Therefore, BIA can be a useful tool to evaluate lipid level.

In this research, BIA data were collected from 87 patients from UMMC. The data processed through statistical analysis to determine the best predictors from the electrical parameters of the SMARTMF device for the lipid profile parameters. However, for LDL-C, no variable could be determined as significant predictor thus an ANN model could not be developed for it. Utilizing the significant predictors, ANN models were developed for TC, TG and HDL-C parameters using 3 different backpropagation algorithms which are the scaled conjugated gradient, Levenberg-Marquardt and resilient. The best performed models are then selected. The best models for the lipid profile parameters are as following, Total Cholesterol-Scaled Conjugate Gradient (TC-SCG) model (Accuracy, 73.3%; Sensitivity, 63.6%; Specificity, 78.9%), Triglyceride-Levenberg-Marquardt (TG-

LM) Model (Accuracy, 76.7%; Sensitivity, 28.6%; Specificity, 91.3%) and High-Density Lipoprotein Cholesterol-Resilient (HDL-SCG) Model (Accuracy, 76.7%; Sensitivity, 95.5%; Specificity, 25.0%)

Future work of research should be performed to improve the classification model for lipid profile. Some of the approaches includes development on larger sample size, optimization of hyperparameters and introduction of new variables or predictors in the model.

Studies have shown that the neural network error rate decreases with the increase of sample size (Raudys & Jain, 1991; Uchimura, Hamamoto, & Tomita, 1995). A bigger sample data is also important to generalize the neural network and enable it to represent the population with better accuracy. Hyperparameters are parameters that directly affects the learning rate, weight and bias of a neural network in a sense of tailoring the details of the solution for the problem at hand. Optimizing the hyperparameters has been linked to improved performance of the algorithms and enable better reproducibility of studies (Feurer & Hutter, 2019).

As of the current study, the variables are only limited to the bioelectrical parameters of the BIA. Based on the T-Test analysis we can observe that height and weight can be a significant variable for the classification model. Furthermore, other variables should be explored such as ethnicity and age group. There are several studies and guidelines that suggests ethnicity as an important risk factor for cholesterol levels (American Heart Association News, 2019; Ismail et al., 2001).

Age on the other hand is a unique predictor of cholesterol levels as it is negatively correlated with LDL-C, but positively correlated with the HDL-C parameter (Kronmal, Cain, Ye, & Omenn, 1993). Therefore, it is important to study the effects of different

variables on cholesterol levels and the impact on its clinical management first and then implement it on a neural network.

In conclusion, this research has shown the potential of both BIA and ANN in the clinical field and how it can be utilized to improve clinical outcome and the general well-being of the population. Further research needs to be carried out to ensure the validity and efficacy of clinical implementation of the techniques.

REFERENCES

American Heart Association News. (2019). Ethnicity a 'risk-enhancing' factor under new cholesterol guidelines. Retrieved from https://www.heart.org/en/news/2019/01/11/ethnicity-a-risk-enhancing-factor-under-new-cholesterol-guidelines

Atiya, A. F. (1990). An unsupervised learning technique for artificial neural networks. *Neural Networks, 3*(6), 707-711. doi:https://doi.org/10.1016/0893-6080(90)90058-S

Cole, K. S., & Cole, R. H. (1941). Dispersion and Absorption in Dielectrics I. Alternating Current Characteristics. *The Journal of Chemical Physics, 9*(4), 341-351. doi:10.1063/1.1750906

Cornish, B. H., Ward, L. C., Thomas, B. J., Jebb, S. A., & Elia, M. (1996). Evaluation of multiple frequency bioelectrical impedance and Cole-Cole analysis for the assessment of body water volumes in healthy humans. *Eur J Clin Nutr, 50*(3), 159-164.

Fabiyi, S. D. (2019). A Review of Unsupervised Artificial Neural Networks with Applications. *International Journal of Computer Applications, 181*, 22-26. doi:10.5120/ijca2019918425

Feurer, M., & Hutter, F. (2019). Hyperparameter Optimization. In F. Hutter, L. Kotthoff, & J. Vanschoren (Eds.), *Automated Machine Learning: Methods, Systems, Challenges* (pp. 3-33). Cham: Springer International Publishing.

Fodor, G. (2008). Primary prevention of CVD: treating dyslipidaemia. *BMJ Clin Evid, 2008*.

Fukunaga, K. (1990). The artificial neural network book.

Galena, H. J. (1992). Complications occurring from diagnostic venipuncture. *The Journal of family practice, 34*(5), 582-584. Retrieved from http://europepmc.org/abstract/MED/1578208

Galushkin, A. I. (2007). *Neural Networks Theory* (1 ed.): Springer-Verlag Berlin Heidelberg.

Gavin, H. (2013). *The Levenberg-Marquardt method for nonlinear least squares curve-fitting problems c ©*.

Gudivaka, R., Schoeller, D. A., Kushner, R. F., & Bolt, M. J. (1999). Single- and multifrequency models for bioelectrical impedance analysis of body water compartments. *J Appl Physiol (1985), 87*(3), 1087-1096. doi:10.1152/jappl.1999.87.3.1087

Gupta, K., & Kang, S. (2011). Implementation of resilient backpropagation & fuzzy clustering based approach for finding fault prone modules in open source software

systems. *International Journal of Research in Engineering and Technology (IJRET), 1*(1), 38-43.

Gutenkunst, R. N., Waterfall, J. J., Casey, F. P., Brown, K. S., Myers, C. R., & Sethna, J. P. (2007). Universally Sloppy Parameter Sensitivities in Systems Biology Models. *PLOS Computational Biology, 3*(10), e189. doi:10.1371/journal.pcbi.0030189

Hansson, G. K. (2005). Inflammation, atherosclerosis, and coronary artery disease. *N Engl J Med, 352*(16), 1685-1695. doi:10.1056/NEJMra043430

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398): John Wiley & Sons.

Ibrahim, F., Faisal, T., Salim, M. I., & Taib, M. N. (2010). Non-invasive diagnosis of risk in dengue patients using bioelectrical impedance analysis and artificial neural network. *Med Biol Eng Comput, 48*(11), 1141-1148. doi:10.1007/s11517-010-0669-z

Ismail, I. S., Nazaimoon, W., Mohamad, W., Letchuman, R., Singaraveloo, M., Hew, F. L., . . . Khalid, B. A. (2001). Ethnicity and glycaemic control are major determinants of diabetic dyslipidaemia in Malaysia. *Diabet Med, 18*(6), 501-508. doi:10.1046/j.1464-5491.2001.00494.x

Jun, M. H., Ku, B., Kim, J., Kim, K. H., & Kim, J. U. (2020). Mediation effect of the duration of diabetes mellitus on the decrease in bioimpedance phase angles in ethnically Korean people: A multicenter clinical study. *Journal of Diabetes Investigation*, 13. doi:10.1111/jdi.13399

Kopin, L., & Lowenstein, C. J. (2017). Dyslipidemia. *Annals of Internal Medicine, 167*(11), 81-96. doi:10.7326/aitc201712050

Krogh, A. (2008). What are artificial neural networks? *Nature Biotechnology, 26*(2), 195-197. doi:10.1038/nbt1386

Kronmal, R. A., Cain, K. C., Ye, Z., & Omenn, G. S. (1993). Total Serum Cholesterol Levels and Mortality Risk as a Function of Age: A Report Based on the Framingham Data. *Archives of Internal Medicine, 153*(9), 1065-1073. doi:10.1001/archinte.1993.00410090025004

Kuziemsky, C. (2016). Decision-making in healthcare as a complex adaptive system. *Healthc Manage Forum, 29*(1), 4-7. doi:10.1177/0840470415614842

Kwon, Y., & Jeong, S. J. (2020). Relative Skeletal Muscle Mass Is an Important Factor in Non-Alcoholic Fatty Liver Disease in Non-Obese Children and Adolescents. *Journal of Clinical Medicine, 9*(10), 9. doi:10.3390/jcm9103355

Kyle, U. G., Bosaeus, I., De Lorenzo, A. D., Deurenberg, P., Elia, M., Gómez, J. M., . . . Pichard, C. (2004). Bioelectrical impedance analysis--part I: review of principles and methods. *Clin Nutr, 23*(5), 1226-1243. doi:10.1016/j.clnu.2004.06.004

Kyle, U. G., Bosaeus, I., De Lorenzo, A. D., Deurenberg, P., Elia, M., Manuel Gómez, J., . . . Pichard, C. (2004). Bioelectrical impedance analysis-part II: utilization in clinical practice. *Clin Nutr, 23*(6), 1430-1453. doi:10.1016/j.clnu.2004.09.012

Lek, S., & Park, Y. S. (2008). Artificial Neural Networks. In (pp. 237-245).

Mallett, S., Halligan, S., Thompson, M., Collins, G. S., & Altman, D. G. (2012). Interpreting diagnostic accuracy studies for patient care. *Bmj, 345*, e3999. doi:10.1136/bmj.e3999

Mat Rifin, H., Robert Lourdes, T. G., Abdul Majid, N. L., Abd Hamid, H. A., Rodzlan Hasani, W. S., Ling, M. Y., . . . Omar, M. A. (2018). Hypercholesterolemia Prevalence, Awareness, Treatment and Control among Adults in Malaysia: The 2015 National Health and Morbidity Survey, Malaysia. *Global Journal of Health Science, 10*(7), 11. doi:10.5539/gjhs.v10n7p11

MIMS. (2020). Dyslipidemia: Diagnosis. Retrieved from https://specialty.mims.com/dyslipidemia/diagnosis

Ministry of Health Malaysia. (2017). Clinical Practice Guideline: Management of Dyslipidaemia. 5th Edition. Retrieved from www.acadmed.org.my

Mohammed, H., Haithem, A., Alani, & George, L. (2015). Comparison between Resilient and Standard Back Propagation Algorithms Efficiency in Pattern Recognition. *International Journal of Scientific and Engineering Research, 6*, 773.

Mohktar, M. S., Ibrahim, F., & Ismail, N. A. (2007, 2007//). *Effects of Abnormal Total Cholesterol Level on Bodycomposition Parameters.* Paper presented at the World Congress on Medical Physics and Biomedical Engineering 2006, Berlin, Heidelberg.

Møller, M. F. (1993). A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks, 6*(4), 525-533. doi:https://doi.org/10.1016/S0893-6080(05)80056-5

National Health Services UK. (2018). Cardiovascular disease. Retrieved from https://www.nhs.uk/conditions/cardiovascular-disease/

Nelles, O. (2013). *Nonlinear system identification: from classical approaches to neural networks and fuzzy models*: Springer Science & Business Media.

Patel, R. V., Peterson, E. L., Silverman, N., & Zarowitz, B. J. (1996). Estimation of total body and extracellular water in post-coronary artery bypass graft surgical patients using single and multiple frequency bioimpedance. *Crit Care Med, 24*(11), 1824-1828. doi:10.1097/00003246-199611000-00011

Pewsner, D., Battaglia, M., Minder, C., Marx, A., Bucher, H. C., & Egger, M. (2004). Ruling a diagnosis in or out with "SpPIn" and "SnNOut": a note of caution. *Bmj, 329*(7459), 209-213. doi:10.1136/bmj.329.7459.209

Portugal, M. R. C., Canella, D. S., Curioni, C. C., Bezerra, F. F., Faerstein, E., Neves, M. F., & Koury, J. C. (2020). Bioelectrical impedance analysis-derived phase angle

is related to risk scores of a first cardiovascular event in adults. *Nutrition, 78*, 7. doi:10.1016/j.nut.2020.110865

Raudys, S. J., & Jain, A. K. (1991). Small sample size effects in statistical pattern recognition: recommendations for practitioners. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 13*(3), 252-264. doi:10.1109/34.75512

Sangi, M., Win, K. T., Shirvani, F., Namazi-Rad, M.-R., & Shukla, N. (2015). Applying a Novel Combination of Techniques to Develop a Predictive Model for Diabetes Complications. *PLOS ONE, 10*(4), e0121569. doi:10.1371/journal.pone.0121569

Schwenk, A., Beisenherz, A., Römer, K., Kremer, G., Salzberger, B., & Elia, M. (2000). Phase angle from bioelectrical impedance analysis remains an independent predictive marker in HIV-infected patients in the era of highly active antiretroviral treatment. *Am J Clin Nutr, 72*(2), 496-501. doi:10.1093/ajcn/72.2.496

Shahid, N., Rappon, T., & Berta, W. (2019). Applications of artificial neural networks in health care organizational decision-making: A scoping review. *PLOS ONE, 14*(2), e0212356. doi:10.1371/journal.pone.0212356

Stoll, G., & Bendszus, M. (2006). Inflammation and atherosclerosis: novel insights into plaque formation and destabilization. *Stroke, 37*(7), 1923-1932. doi:10.1161/01.Str.0000226901.34927.10

Süt, N., & Şenocak, M. (2007). Assessment of the performances of multilayer perceptron neural networks in comparison with recurrent neural networks and two statistical methods for diagnosing coronary artery disease. *Expert Systems, 24*(3), 131-142. doi:https://doi.org/10.1111/j.1468-0394.2007.00425.x

Toso, S., Piccoli, A., Gusella, M., Menon, D., Bononi, A., Crepaldi, G., & Ferrazzi, E. (2000). Altered tissue electric properties in lung cancer patients as detected by bioelectric impedance vector analysis. *Nutrition, 16*(2), 120-124. doi:10.1016/s0899-9007(99)00230-0

Trevethan, R. (2017). Sensitivity, Specificity, and Predictive Values: Foundations, Pliabilities, and Pitfalls in Research and Practice. *Frontiers in public health, 5*, 307-307. doi:10.3389/fpubh.2017.00307

Uchimura, S., Hamamoto, Y., & Tomita, S. (1995, 27 Nov.-1 Dec. 1995). *Effects of the sample size in artificial neural network classifier design.* Paper presented at the Proceedings of ICNN'95 - International Conference on Neural Networks.

Urbina, E. M., & Daniels, S. R. (2008). Chapter 14 - Hyperlipidemia. In G. B. Slap (Ed.), *Adolescent Medicine* (pp. 90-96). Philadelphia: Mosby.

Vrbaški, M., Doroslovački, R., Kupusinac, A., Edita, S., & Ivetic, D. (2019). Lipid profile prediction based on artificial neural networks. *Journal of Ambient Intelligence and Humanized Computing*. doi:10.1007/s12652-019-01374-3

Walczak, S., & Cerpa, N. (2003). Artificial Neural Networks. In R. A. Meyers (Ed.), *Encyclopedia of Physical Science and Technology (Third Edition)* (pp. 631-645). New York: Academic Press.

Warnick, G. R., Kimberly, M. M., Waymack, P. P., Leary, E. T., & Myers, G. L. (2008). Standardization of Measurements for Cholesterol, Triglycerides, and Major Lipoproteins. *Laboratory Medicine, 39*(8), 481-490. doi:10.1309/6ul9rhjh1jffu4py

World Health Organization. (2017). Cardiovascular diseases. Retrieved from https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

World Health Organization. (2018). Noncommunicable diseases country profiles 2018. Retrieved from https://www.who.int/nmh/publications/ncd-profiles-2018/en/