ANALYSIS OF INFLAMMATION, METABOLIC AND CLINICAL MARKERS IN PREDICTING FAT MASS IN HIV-POSITIVE MALES USING ARTIFICIAL NEURAL NETWORK

NURUL FARHAH SHAMSUDDIN

FACULTY OF ENGINEERING
UNIVERSITY OF MALAYA
KUALA LUMPUR

2021

Universiti Malaya

ANALYSIS OF INFLAMMATION, METABOLIC AND CLINICAL MARKERS IN PREDICTING FAT MASS IN HIV-POSITIVE MALES USING ARTIFICIAL NEURAL NETWORK

NURUL FARHAH SHAMSUDDIN

DISSERTATION SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF ENGINEERING SCIENCE

FACULTY OF ENGINEERING
UNIVERSITY OF MALAYA
KUALA LUMPUR

2021

Universiti Malaya

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Nurul Farhah Shamsuddin

Matric No: 17036246/3

Name of Degree: Master of Engineering Science

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

Analysis of Inflammation, Metabolic and Clinical Markers in Predicting Fat Mass in Hiv-Positive Males Using Artificial Neural Network

Field of Study: Machine Learning

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:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name:

Designation:

University of Malaya

ANALYSIS OF INFLAMMATION, METABOLIC AND CLINICAL MARKERS IN PREDICTING FAT MASS IN HIV-POSITIVE MALES USING ARTIFICIAL NEURAL NETWORK

ABSTRACT

Increased inflammation has been discovered in people living with Human Immunodeficiency Virus (HIV), which has been associated with the development and advancement of metabolic diseases. Immune activation and inflammation have been revealed to affect the redistribution of fat in people living with HIV (PLWH) whether the individual is on treatment with antiretroviral therapy (ART) or treatment naïve. Therefore, the assessment of body fat composition of PLWH based on markers of inflammation is important to develop keen diagnosis and prognosis of medical conditions facing by these patients, for example lipodystrophy syndrome, metabolic syndrome, dyslipidemia and diabetes mellitus. The purpose of this study is to investigate the relationship among inflammation marker, interleukin-6 (IL-6), clinical and metabolic factors with fat mass (FM) in males living with HIV (MLWH) using artificial neural network (ANN) and to assess the contribution of metabolic and clinical factors alongside inflammation marker, IL-6 to body fat composition in MLWH. Five sets of ANN models were constructed, and each set of network model was produced and manipulated based on inclusion of inflammation marker, IL-6 as an independent variable and on statistically significant association of independent variables with dependent variable, FM. First set of ANN models used inflammation, metabolic and clinical variables as input, the second set of ANN models used metabolic and clinical variables as input, the third set of ANN models used body mass index (BMI) and waist-hip-ratio (WHR) as input, the fourth and fifth set of ANN models used statistically significant parameters associated with FM but each set applied different training split ratio. Dependent variable consisted of body

Universiti Malaya

composition variable, FM collected from whole-body DEXA scan was selected as output. Comparison of model performance were assessed through the model error function, error sum of squares (SSE), relative error (RE) and mean predictive accuracy percentage (MPA%). The MPA% obtained by Set A was 84.84% with ten hidden nodes in its single-hidden layer ANN, 84.02% with ten hidden nodes for Set B, 80.09% with eight hidden nodes for Set C, 85.26% with four hidden nodes for Set D and 85.48% with two hidden nodes for Set E. Through independent sample *t*-test conducted on Set A and Set B ANN performance to discover the influence of the inflammation marker, IL-6, no significant difference was found in the MPA% between ANN models using clinical, metabolic and inflammation data and ANN models using clinical and metabolic variables, $t\,(13) = 0.75$, $p < .05$, 95% C.I. [-0.95% – 1.96%]. However, statistically significant difference in prediction accuracy was found on fat mass in MLWH at the $p < .05$ level for the five ANN models sets [$F\,(4, 45) = 8.802$, $p = 0.00$]. The findings show that ANN technique was able to triangulate the relationship between body fat composition and clinical, metabolic and inflammation data of MLWH even though the addition of inflammation marker, IL-6 did not significantly improve the ANN performance.

Keywords: artificial neural network (ANN); body composition; HIV; inflammation; metabolism.

Universiti Malaya

ABSTRAK

Peningkatan keradangan telah ditemui pada individu HIV-positif yang telah dikenalpasti sebagai penyebab dalam peningkatan penyakit metabolik. Pengaktifan imun dan keradangan telah didedahkan mempengaruhi pengagihan semula lemak dalam pesakit HIV tidak kira sama ada pesakit itu menjalani rejimen antiretroviral atau tidak. Oleh itu, penilaian komposisi lemak badan pesakit HIV berdasarkan faktor keradangan adalah perlu untuk pembangunan diagnosis dan prognosis penyakit bagi pesakit-pesakit ini seperti sindrom lipodistrofi, sindrom metabolik, dislipidemia dan kencing manis. Tujuan kajian ini adalah untuk mengkaji hubungan antara penanda keradangan, interleukin-6 (IL-6), faktor metabolik dan klinikal dengan jisim lemak dalam pesakit HIV lelaki menggunakan rangkaian saraf buatan dan untuk menilai sumbangan faktor risiko metabolik dan klinikal dengan penanda keradangan, IL-6 pada komposisi lemak badan dalam pesakit HIV lelaki. Lima set model rangkaian saraf buatan telah dibina dan setiap set model rangkaian saraf dihasilkan dan dimanipulasi berdasarkan kehadiran penanda keradangan, IL-6 sebagai pembolehubah bebas. Set pertama model (Set A) menggunakan faktor keradangan, metabolik dan klinikal sebagai input, set kedua model (Set B) menggunakan faktor metabolik dan klinikal sebagai input, set ketiga model (Set C) menggunakan indeks jisim badan dan nisbah pinggang kepada pinggul sebagai input dan set keempat model (Set D) dan set kelima model (Set E) menggunakan parameter yang signifikan secara statistik dengan jisim lemak tetapi kedua-dua model berbeza dalam penentuan kadar nisbah latihan. Pembolehubah bergerak balas adalah pembolehubah komposisi badan, jisim lemak yang diambil daripada imbasan DEXA yang mana ia dipilih sebagai faktor output. Perbandingan prestasi setiap model dikira melalui fungsi ralat model iaitu kesilapan jumlah kuadrat dan ralat relatif serta peratusan ketepatan ramalan. Purata peratusan ramalan diperolehi Set A adalah 84.84% dengan sepuluh nod tersembunyi dalam rangkaian saraf satu lapisan, 84.02% dengan sepuluh nod

Universiti Malaya

tersembunyi untuk Set B, 80.09% dengan lapan nod tersembunyi untuk Set C, 85.26% dengan empat nod tersembunyi untuk Set D dan 85.48% dengan dua nod tersembunyi untuk Set E. Melalui ujian t-sampel bebas yang dijalankan ke atas prestasi rangkaian saraf buatan Set A dan Set B untuk mengenalpasti pengaruh IL-6, didapati tiada perbezaan yang signifikan yang dilaporkan dalam peratusan purata ramalan antara kedua-dua set, t (13) = 0.75, p< .05, 95% C.I. [-0.95% – 1.96%]. Tetapi, hasil kajian statistik mendapati terdapat perbezaan yang signifikan dalam ramalan ketepatan jisim lemak untuk kelima-lima set model iaitu [F (4,45) = 8.802, p = 0.00] pada p <.05. Hasil kajian menunjukkan bahawa teknik rangkaian saraf buatan dapat menjumpai hubungankait antara komposisi lemak badan dan data klinikal, metabolik dan keradangan pesakit lelaki HIV walaupun sumbangan penanda keradangan, interleukin-6 tidak jelas dalam meningkatkan prestasi rangkaian saraf buatan.

Kata kunci: rangkaian saraf buatan; komposisi tubuh badan; HIV; keradangan; metabolisma.

Universiti Malaya

ACKNOWLEDGEMENTS

It is a pleasure to thank several people who contributed to the preparation of this thesis.

I would like to thank my three supportive supervisors, Ir. Dr. Mas Sahidayana Mokhtar, Assoc. Prof. Dr. Reena Rajasuriar and Dr Wan Safwani Wan Kamarul Zaman for their patience and guidance. Their thorough review and comments proved invaluable to the completion of this thesis.

I wish to thank laboratory staffs from Centre for Excellence for Research in AIDS (CERIA), University of Malaya who provided the database for the analysis, for their input, support and assistance in addition to my colleagues at Biomedical Engineering Department, Faculty of Engineering, University of Malaya for their input and peer support.

My sincere appreciation to my husband who always at my side, for his constant encouragement particularly during our hard time.

Lastly, I wish to thank my parents for their never-ending care and great love bestowed to me.

Universiti Malaya

TABLE OF CONTENTS

Universiti Malaya

Universiti Malaya

Universiti Malaya

LIST OF FIGURES

Universiti Malaya

Universiti Malaya

LIST OF TABLES

Universiti Malaya

LIST OF SYMBOLS AND ABBREVIATIONS

AIDS	:	Acquired immunodeficiency syndrome
ANN	:	Artificial neural network
BMI	:	Body mass index
cART	:	Combination of antiretroviral therapy
CCR5	:	C-C chemokine receptor type 5
CD4	:	Cluster of differentiation 4
CD8	:	Cluster of differentiation 8
CPRs	:	Clinical prediction rules
CRP	:	C-reactive protein
CVD	:	Cardiovascular disease
DEXA	:	Dual-energy x-ray absorptiometry
FFM	:	Fat free mass
FM	:	Fat mass
HDL	:	High-density lipoprotein
HIV	:	Human immunodeficiency virus
IL-6	:	Interleukin-6
IL-8	:	Interleukin-8
IL-Iβ	:	Interleukin-1 beta
LDL	:	Low-density lipoprotein
MLWH	:	Males living with HIV
MPA%	:	Mean predictive accuracy percentage
NNRTI	:	Non-nucleoside reverse transcriptase inhibitors
NRTI	:	Nucleoside reverse transcriptase inhibitors
PI	:	Protease inhibitor
PLWH	:	People living with HIV
RE	:	Relative error
RNA	:	Ribonucleic acid
RT	:	Reverse transcriptase
SD	:	standard deviation
SSE	:	Error sum of squares
T2DM	:	Type-2 diabetes mellitus
TNF-α	:	Tumor necrosis factor alpha
WHO	:	World health organization
WHR	:	Waist-hip ratio
WLWH	:	Women living with HIV

Universiti Malaya

CHAPTER 1: INTRODUCTION

1.1 Background of Study

People living with HIV (PLWH) infection have long suffered from a high prevalence of mortality and morbidity compared to the uninfected individuals in the absence of antiretroviral therapy. Human immunodeficiency virus or HIV is a group of retroviruses that interferes and alters the immune system by attacking immune T-cells known as cluster of differentiation 4 (CD4) cells (Sarngadharan et al., 1985). The progressive loss of CD4 T-cells eventually leads to the collapse of the immune system and the development of acquired immunodeficiency syndrome (AIDS), the latter stage in HIV infection (Patel et al., 2014). If untreated, patients with AIDS will die from consequences of a compromised immune system. Researchers and physicians have formulated antiretroviral medications to treat HIV infection. PLWH are required to consume a combination of antiretroviral therapy (cART) to effectively suppress viral replication and allow the reconstitution of the immune system (Anglemyer et al., 2014). cART has showed significant success when it is introduced and effectively reduces the rate of morbidity and mortality for PLWH, hence enable them to lead a near normal lifespans as their non-infected peers (Fauci & Folkers, 2012).

However, some cART drugs have been associated with the development of comorbidities such as cardiovascular disease (CVD) (Nsagha et al., 2015), type-2 diabetes mellitus (T2DM) (Nansseu et al., 2018), metabolic syndrome (Nguyen et al., 2016) and lipodystrophy syndrome (Finkelstein et al., 2015). The physiology and pathology of these adverse events relationship with HIV and cART are still not fully understood and various studies (Grant et al., 2016; John et al., 2001; Ross et al., 2008) have been conducted to investigate this. One of the potential explanations is that increased inflammation and immune activation experienced by PLWH affects metabolism and becomes a risk factor for lipodystrophy syndrome, CVD, T2DM and overall mortality for the population

Universiti Malaya

(Aounallah et al., 2016). HIV infection is associated with chronic immune activation, and HIV-infected individuals are characterized by higher inflammatory levels than non-HIV-infected individuals (Aberg, 2012).

Alteration of fat in PLWH especially fat associated with subcutaneous adipose tissues has largely been linked to increase inflammation caused by the activation of immune cells and cytokines (Funderburg & Mehta, 2016). The number of macrophages in immune system are increased and the levels of tumor necrosis factor alpha (TNF-α), interkeukin-6 (IL-6) and interleukin-8 (IL-8) are detected to be high in HIV-positive individuals with lipodystrophy (Caron-Debarle et al., 2010). It marks the inflammation as a potential risk for fat abnormalities in PLWH. The condition is further worsened with long term infection and prolonged consumption of cART medication (Villarroya et al., 2010).

Predicting medical conditions for PLWH especially the underlying adverse conditions constitute large effort and determination from researchers and physicians alike. With timely and informative action, medical intervention could present as a rewarding undertaking in combating many side-effects of cART. It is well-known that specific antiretroviral drug classes such as protease inhibitor (PI) (Ucciferri et al., 2013) are associated with metabolic adverse events despite its overwhelming success in suppressing viral replication and preventing the development of AIDS defining illnesses.

Prediction and diagnosis of medical conditions has its own regulation and rule which researchers, physicians and clinicians must comply according to the Clinical Prediction Rules (CPRs). CPRs are a field of medical research in which rules for prognostics or diagnoses of disease are developed by measuring a combination of medical signs, symptoms and other physiological and clinical findings (McGinn et al., 2008). Algorithm is used to determine the scores by using mainly linear regression or similar linear statistical methods (Beraldo et al., 2015; Brown et al., 2013). By following the proposed

Universiti Malaya

set of rules and guidelines, physicians could advance and thus, prevent the morbidity and mortality of said disease. However, decrease in accuracy and generalizability are preventing the clinicians to apply CPRs in their practice(Adams & Leveson, 2012). The reason found for lower accuracy includes differences from the initial and latter patient population under treatment where the CPR is not extensible and less adaptive in adjusting to the difference.

Engineering approaches has been considered and studied extensively as a new method to prediction in prognostics and health management field. Among its adaptive mathematical methods is, Artificial Neural Network (ANN) which has become popular recently as a compelling prediction tool. The main reasons are it is able to model non-linear relationships, handle adaptive learning, pattern recognition and classification, features regarded as important in building medical predictive model profiles.

1.2 Study Overview

To achieve the study aims, this thesis utilized non-linear modelling analysis that incorporated feed-forward neural network with back propagation technique in order to enable non-linear mapping between the input and output and represent the complex function with ease. This study was a collaborative work between Biomedical Engineering Department of University Malaya and Centre of Excellence for Research in AIDS (CERIA), Department of University Malaya where the data used for analysis were part of a larger project to explore the biological factors associated with aging in PLWH. The data used consisted of two parts; the first part of the data comprised of clinical / demographic data, metabolic data and inflammation marker data while the second part of the database was obtained from iPacs system database of University Malaya Medical Centre which stores the body composition data measured using dual-energy x-ray

Universiti Malaya

absorptiometry (DEXA). Number of participants involved in this study were 71 males living with HIV (MLWH). The study seeks to better understand the relationship between fat mass (FM) with the marker of inflammation, IL-6, metabolic and clinical markers including body mass index (BMI) and waist-hip ratio (WHR) in MLWH.

The first aim of this study was to develop models to predict the relationship between FM with the marker of inflammation, IL-6, metabolic and clinical markers in MLWH using ANN. Five set of ANN models were developed using single-hidden layer multilayer perceptron neural network. Each set contained ten models with different number of hidden nodes where each set of network models was produced and manipulated based on inclusion of inflammation marker IL-6 as independent variables and on statistically significant association of independent variables with dependent variable, FM. This approach was taken to measure the function error and prediction accuracy for every model.

To achieve the second aim of this study, the developed ANN models in the first objective were compared to assess the minimum function error of each sets and to evaluate the prediction accuracy of FM with the marker of inflammation, IL-6, metabolic and clinical markers. Zero or almost zero function error is desired for the ANN models because low error signifies the ANN model is well-developed and shows good performance. Therefore, evaluation of minimum function error is important so that the model performance can be assessed, and new adjustment can be introduced. Whereas the prediction accuracy will show the ability of ANN model to predict body FM of MLWH by using ANN.

Furthermore, third aim of study was developed to determine whether the BMI and WHR are factors influencing the fat composition in MLWH. It is important to assess the

Universiti Malaya

anthropometric measure as a form of simpler approach to prediction performance of FM within MLWH.

Descriptive statistics of clinical, metabolic, inflammation and body FM data were completed to observe the trend within MLWH cohort. In addition, bivariate linear regression analyses were conducted between clinical, metabolic and inflammation data as predictor variable and FM as criterion variable. The analysis was completed to test simple hypotheses of association between the variables.

1.3 Motivation of Research

cART has led to an improvement of health in PLWH and prevented a high rate of morbidity and mortality previously found in untreated disease. It was estimated that over 37.9 million people globally were living with HIV by the end of 2018 and 23.3 million or 62% of PLWH were receiving antiretroviral treatment in 2018 (UNAIDS, 2019). However, specific classes of cART have been associated with undesirable side effects, comprising of short-term side effects and long-term side effects. For example, muscle wasting continues to be a problem for many patients (Nansseu et al., 2018) and the use of cART brought about new concerns related to nutritional status such as the complication of metabolic syndrome (Pontes Signorini et al., 2012) and lipodystrophy syndrome (Finkelstein et al., 2015) in PLWH receiving the treatment. Many of the adverse events are manageable.

Other possible adverse events which are associated with cART regimen consumption apart from metabolic syndrome and lipodystrophy syndrome includes among others dyslipidemia (Tripathi et al., 2013). These afflictions commonly referred as metabolic disorders where the metabolism process fail to function accordingly or in severe state causing the associated organs to become diseased or function abnormally. The disruption

Universiti Malaya

of normal metabolic process is caused by the risk factors associated with metabolic syndrome, dyslipidemia, insulin resistance and T2DM where these risk factors contribute to cellular dysfunction and redox imbalance which promote the progression of pro-oxidative environment. The oxidative stress further lead to damage biomolecules, which are highly reactive in nature and can promote cell and tissue dysfunction leading to metabolic diseases (Rani et al., 2016). Metabolic disorders are the subject of ongoing research due to its intertwined and complex structure caused by various factors such as glucose intolerance and the overabundance of certain proteins (Lake & Currier, 2013).

The inflammatory component of metabolic distortion has been increasingly recognized as one of the risk factors aside from the central obesity and insulin resistance which are thought to represent the common underlying factors of the syndrome (Bourgi et al., 2018). Measurement of markers of inflammation has been proposed as a method to identify metabolic syndrome where many studies have found a close correlation with the syndrome and dyslipidemia, abdominal obesity, low insulin sensitivity and hypertension with increasing level of inflammatory markers (Funderburg & Mehta, 2016).

The countermeasures taken to overcome the side-effects induced by the metabolic syndrome are regular exercise, addressing nutrition intake (Botros et al., 2012), use of ART drugs with more neutral metabolic profiles (McComsey et al., 2011) and cosmetic treatment for adipose tissue dysfunction (Shuck et al., 2013) have been suggested and applied with different level of success. The weaknesses of the suggested methods include unclear measure of short-term and long-term benefit, delayed and interruption of antiretroviral therapy treatment and modest improvements in the signs of lipodystrophy measured by radiological methods.

The rise in data learning technology has provided new approach in dealing with shortcoming of past measures. With frequent and accurate data collection from lab test

Universiti Malaya

results and diagnostic tests, it gives researchers and clinicians a new and diverse analytical tool in diagnosing and predicting the changes caused by metabolic diseases. Early detection of the medical complication permit physician to apply appropriate interventions and prevent the adverse outcome. It is essential to push for further research and development to expand predictive medical analytics into robust, practical and everyday standard medical tools. ANN can be a promising tool since it able to model non-linear data relationships and quickly adapt to new sets of data, hence, providing a computational model suitable to specific patient demographic and needs.

1.4 Research Objectives

1) To develop ANN models towards predicting the relationship between FM and IL-6, metabolic and clinical markers in MLWH.
2) To evaluate the prediction accuracy of FM with IL-6, metabolic and clinical markers in MLWH.
3) To determine whether the BMI and WHR are factors influencing the fat composition in MLWH.

1.5 Research Questions

1) What is the contribution of inflammation marker, IL-6 to the accuracy of models predicting fat composition in MLWH patients?
2) What is the use of clinical and metabolic factors in predicting FM in MLWH patients?
3) Does using BMI and WHR as variable input contribute to the accuracy of predicting fat composition in MLWH patients?

Universiti Malaya

1.6 Scope of Work

The work scope of the thesis are as follows:

1) Data collection (consisting of clinical, metabolic, inflammation and body composition data extracted from Centre of Excellence for Research in AIDS (CERIA), University Malaya database and iPacs system database of University Malaya Medical Centre.)
2) Data pre-processing (Preparation and cleaning of data by removing outliers, missing data, irrelevant data and formatting.)
3) Data description (summary statistics summarizing values and measures in sample data.)
4) Bivariate linear regression analysis.
5) Analysis of relationship between inflammation marker, IL-6, metabolic and clinical parameters with FM in MLWH patients using ANN – multilayer perceptron neural network.
6) Comparison of ANN models constructed using inflammation, metabolic and clinical data as input variables, ANN models using only metabolic and clinical data as input variables, ANN models using BMI and WHR as input variables and ANN models using statically significant parameters associated with FM.

1.7 Significance of Research

It is necessary to conduct the study as it can be a new form and complimentary diagnosis to lower the potential risk factor involved in progression of CVD and other fatal diseases by discovering the role of inflammation in the measurement of FM in MLWH. This finding in turn can offer and provide different perspective on how inflammation and immune activation contribute to the alteration and modification of fat in MLWH patients

Universiti Malaya

especially in individuals who have been on cART for a long time. The study will also contribute to previous research done on the role of inflammation and its association with fat distribution specifically lipid, cholesterol, low-density lipoprotein (LDL) and high-density lipoprotein (HDL) abnormalities in HIV infection in MLWH.

It is also required to investigate the non-linear relationship between inflammation and FM in MLWH and this study uses deep learning method based on ANN to investigate the pattern recognition and solve the problem which common method such as linear regression is unable to due to its fixed, rigid linear model and deemed as unsuitable for data with nonlinearities tendency with certain assumptions must be met. ANN is preferred as it is known to identify complex relationships even though the data are intricate, ill-defined and ill-structure in nature and build accurate model with the data it is trained and tested. This in turn could be used to detect imperceptible pattern and improve the prediction power of the diagnosis of metabolic abnormalities in HIV infection which could lead to preventive medicine for MLWH patients. This research is a novel research and have not been studied yet by any other researchers. This study is the first research trying to analyze the inflammation, metabolic and clinical markers in predicting FM in MLWH using ANN.

1.8 Organization of Thesis

This thesis is divided into six chapters. Chapter 1 outlines the study background, overview of study, motivation of research, research objectives, research questions, scope of work and significance of the research conducted. Chapter 2 covers literature review in the field of HIV and how metabolic-related diseases are associated with HIV. Review on non-linear modelling method employing ANN is also discussed together with its contribution in clinical field. Chapter 3 recounts the methodology design involved in this

Universiti Malaya

study besides the data collection itself. Chapter 4 presents the result acquired from the analysis. The results include the subject characteristics, linear regression analysis and artificial neural network analysis. Chapter 5 discusses the findings obtained from the analysis and evaluate the result according to the purpose of the study. Chapter 6 covers the summary of this thesis including the conclusions drawn from the research, the limitation and suggestion of ideas for related future work. Concluding this thesis is the bibliography.

Universiti Malaya

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

This chapter introduces the HIV and cART regimen undertook by PLWH. Metabolic disorders suffered by these individuals are also reviewed including the cART regimen effect on these illnesses. Besides that, body composition of PLWH and the assessment method – DEXA are also reviewed together with how inflammation affects the body composition of these individuals. Prediction method used in this study, ANN is also reviewed and the past success of HIV research employing ANN as their technique is also discussed. In addition, comparison of conventional statistical method and ANN was also described.

2.2 Human Immunodeficiency Virus (HIV) and Combination of Antiretroviral Therapy (cART)

In 1981, five cases of Pneumocystis carinii pneumonia were presented to the United States Centre for Disease Control in homosexual men who had also presented with decreased CD4 T cell count resulting in a deficit in cell-mediated immunity (Gottlieb et al., 1981). The men were later found to be suffering from an acquired immune deficiency syndrome (AIDS), caused by a retrovirus named human T-lymphotropic virus type III / lymphadenopathy-associated virus (Sarngadharan et al., 1985), or as it called nowadays, human immunodeficiency virus (HIV). As many as 74.9 million people have become infected with HIV and 32.0 million people have died from AIDS-related illnesses since the start of the epidemic until end of 2018 (UNAIDS, 2019).

The retrovirus attacks CD4 T-helper cells, a type of white blood cell in the immune system and replicates persistently inside these cells which later weakened the body line of defense system gradually (Klatzmann et al., 1984). HIV infection, when left untreated,

Universiti Malaya

eventually progress to CD4 T-cell deficiency that is associated with AIDS-related morbidities and death (Figure 2.1) (Palella Jr et al., 1998). These morbid events include infections with *pneumocystis jiroveci* causing pneumonia (Huang et al., 2017), mycobacterium tuberculosis, *M. avium* intracellular (Bruchfeld et al., 2015) and herpes simplex virus (Desai & Kulkarni, 2015) among others. PLWH are also at increased risk for developing Kaposi's sarcoma, non-Hodgkin lymphoma and other cancers (Yanik et al., 2016). The rate in which disease progresses from primary infection to AIDS-related morbidity is variable among PLWH. For instance, there are typical progressors who develop AIDS-defining illnesses following an extended period of time and rapid progressors who rapidly progress to AIDS-defining illnesses within 2-3 years following primary infection (Poorolajal et al., 2016). The rate of disease progression can be evaluated by immunological and virologic markers. Peripheral blood CD4 T-cell count is a well-characterized predictor of disease progression (Doitsh & Greene, 2016). Plasma HIV RNA level also serves as an important predictor of disease progression (Rouzioux & Avettand-Fenoël, 2018).

cART treatment helps combat the infection and restores CD4 T-cell numbers and function (Maartens et al., 2014). cART usually comprises of a combination of at least three antiretroviral drugs from at least two different antiretroviral classes and this combination has been shown to be effective in achieving sustained suppression of HIV replication thus allowing CD4 T-cells to reconstitute and the return to health for PLWH (Teeraananchai et al., 2017). These drugs include PI and nucleoside reverse transcriptase inhibitors (NRTI) and were initiated in 1996 as a new standard of care for PLWH. These drugs along with others can inhibit HIV replication in multiple steps (Figure 2.2) and were found to markedly suppress the viral load below detection limit.

Universiti Malaya

There are six groups of antiretroviral agents available in Malaysia as shown in Table 2.1 and the antiretroviral drugs differ from each other depending on their antiretroviral classes (Medical Development Division, 2011). Each of the drugs has its own mode of action and outcome based on its specifically designed chemical structure in the viral lifecycle (Figure 2.2). NRTI interact with the substrate binding site of HIV reverse transcriptase (RT) (Kakuda, 2000). Non-nucleoside reverse transcriptase inhibitors (NNRTI) differ from NRTI in that they bind specifically with a non-substrate binding site of the RT enzyme, disrupting the enzymes catalytic site. PI inhibit the protease enzyme, thus preventing the cell from cleaving the proteins into active viral particles (Wynn et al., 2004). Fusion inhibitors block the attachment, co-receptor binding and fusion of the viral particle and prevent viral capsid entry into the host cell (Greenberg et al., 2004). C-C chemokine receptor type 5 (CCR5) inhibitors inhibit CCR5 signaling, are antagonists of the CCR5 receptor and prevent viral entry into the host cell (Dorr et al., 2005). Integrase inhibitors are the newest class of antiretroviral therapy and work by inhibiting the insertion of the HIV-1 pro-viral DNA into the host cell genome (De Clercq & Herdewijn, 2010).

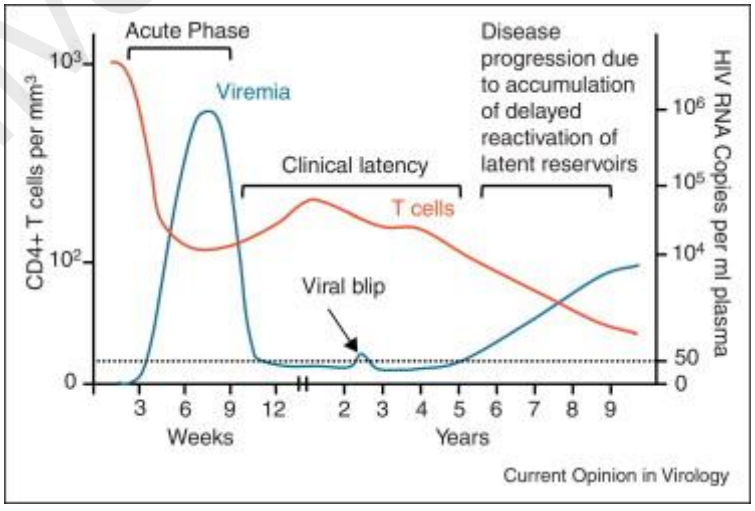

Figure 2-1 CD4 T-cell loss and viral load progression in HIV infection (adapted from "Mathematical Model of Viral Latency" authored by Christian Selinger and Michael G. Katze and published by Current Opinion in Virology (July 2013) (Selinger & Katze, 2013))

Universiti Malaya

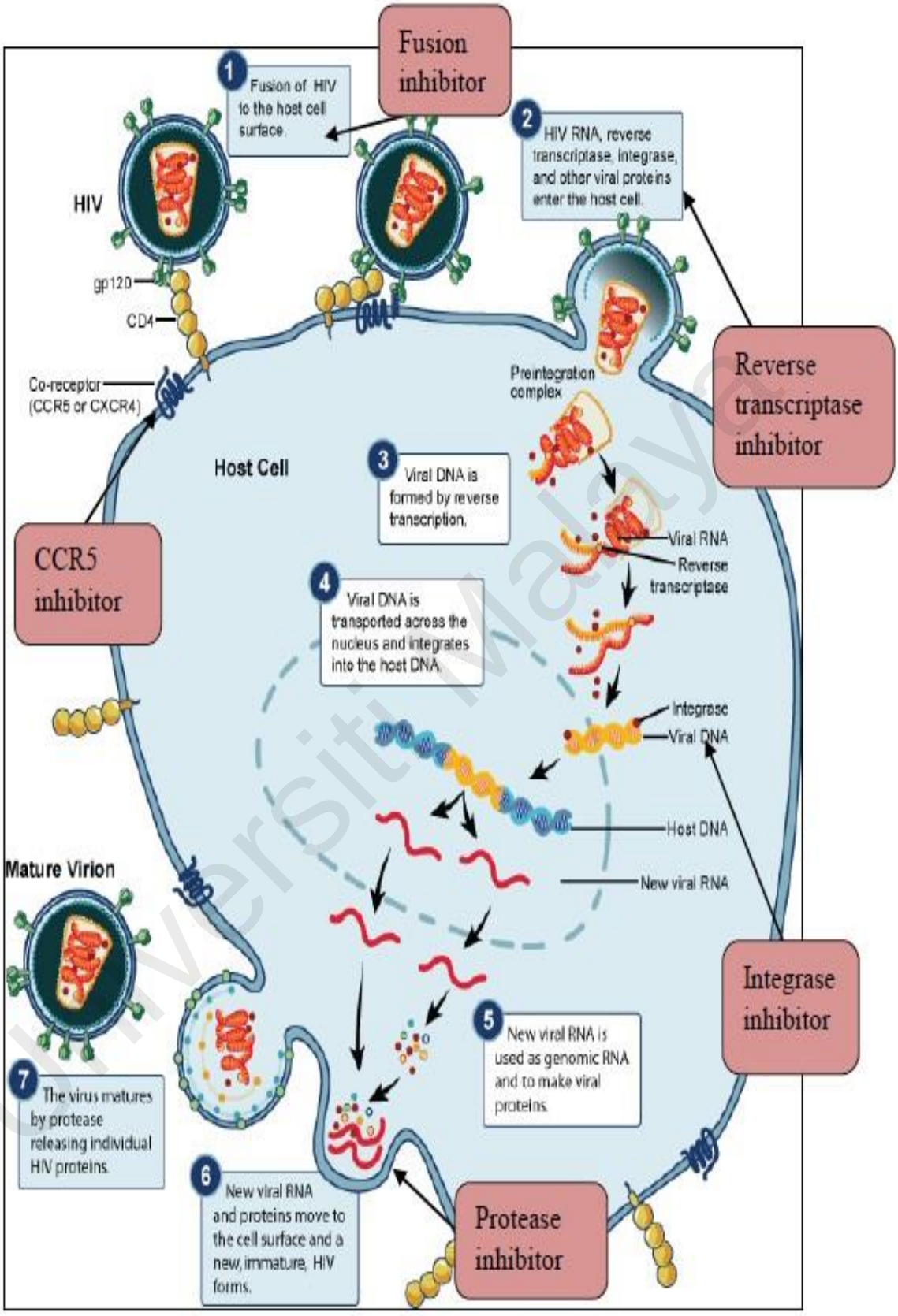

Figure 2-2 Mode of action of antiretroviral drugs (adapted from (*HIV Replication Cycle*, 2018))

Universiti Malaya

Table 2-1 Antiretroviral drugs currently licensed in Malaysia (adapted from Guidelines for the Management of Adult HIV Infection with Antiretroviral Therapy, Medical Development Division, Ministry of Health Malaysia, December 2011 (Medical Development Division, 2011))

1	**Nucleoside Reverse Transcriptase Inhibitors (NRTI) / Nucleotide Reverse Transcriptase Inhibitors (NtRTI)**
	Zidovudine (AZT) Didanosine buffered (ddl) or enteric coated (ddl EC) Stavudine (d4T) Lamivudine (3TC) Abacavir (ABC) Tenofovir (TDF) Emtricitabine (FTC) (available as combination with tenofovir)
2	**Non-Nucleoside Reverse Transcriptase Inhibitor (NNRTI)**
	Nevirapine (NVP) Efavirenz (EFV) Etravirine (Intelence)
3	**Protease Inhibitors (PI)**
	Indinavir (IDV) Lopinavir/ritonavir (Kaletra) Saquinavir Atazanavir (ATV) Darunavir (Prezista) Ritonavir Nelfi navir (Viracept)
4	**Integrase Inhibitors**
	Raltegravir (Isentress)
5	**CCR5 Antagonists**
	Maraviroc (Calsentri)
6	**Fusion Inhibitor**
	Enfuritide (Fuzeon)

2.3 Metabolic Disorders Faced by People Living with HIV (PLWH)

Metabolic disorder is defined by irregularity in metabolism process caused by unregulated and abnormal chemical reactions directly and indirectly linked to body metabolism. Common diseases associated with metabolic irregularity are lipodystrophy, metabolic syndrome, dyslipidemia and T2DM.

Universiti Malaya

2.3.1 Lipodystrophy Syndrome

Lipodystrophy is a condition where PLWH have their body fat changes including either build-up or loss of body fat or both. Lipodystrophy can be divided into lipoatrophy and lipohypertrophy. Lipoatrophy is associated with localized loss of fat tissue or peripheral fat wasting. It is classically noticeable around the face (facial lipoatrophy), arms, legs and buttocks (peripheral lipoatrophy) (Hammond et al., 2010) among PLWH as shown in Figure 2.3 (b) and (d). The preservation of body lean mass (primarily muscle) in lipoatrophy distinguishes this condition from HIV wasting syndrome, which is defined as the generalized loss of body fat and body lean mass with an involuntary weight loss > 10% of baseline body weight (Lichtenstein, 2005).

Lipohypertrophy is defined by the accumulation of adipose tissue or total body fat in certain regions, mostly in the head and neck region, trunk, dorsocervical and intraabdominal regions as shown in Figure 2.3 (a) and (c) (Leung & Glesby, 2011). It is worthy to be noted that there is also fat accumulation which leads to breast enlargement in both male and females and parotid enlargement though this is less frequent (Alencastro et al., 2017). Women living with HIV (WLWH) have also been reported to have high possibility to incur combined peripheral and central lipoatrophy double that of normal non-infected women after adjustment for age and race. The research did not however, investigate the incidence in MLWH (Tien et al., 2003).

Universiti Malaya

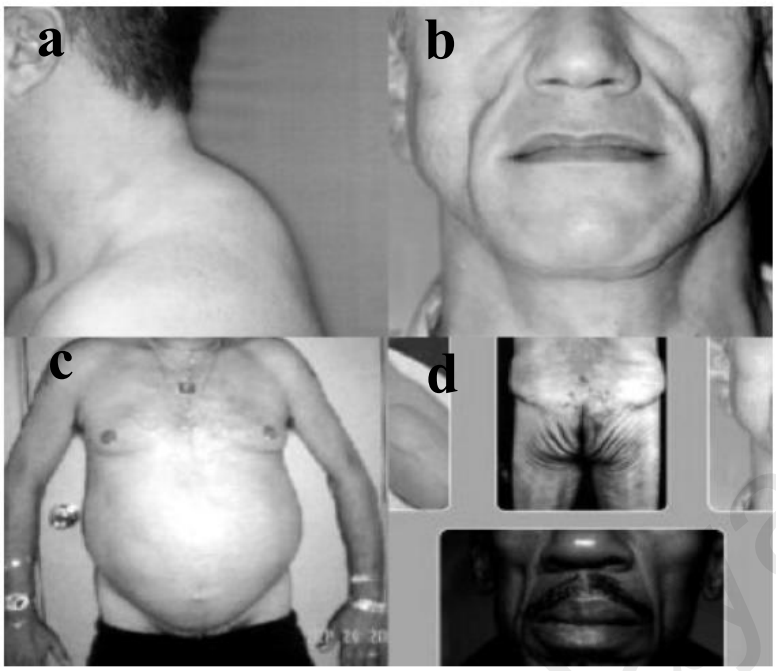

Figure 2-3 Image on top left (a) shows HIV-related dorsocervical fat pad, the so-called "buffalo hump; image on top right (b) shows HIV-related facial lipoatrophy; image at bottom left (c) shows HIV-related abdominal lipohypertrophy; image at bottom right (d) shows HIV-related bottom and facial lipoatrophy (adapted from (Vergel, 2008))

Risk factors associated with the HIV-associated lipoatrophy include therapy with thymidine analogue NRTIs even though initially PI was suspected to be the culprit. Other factors associated to lipoatrophy are older age, low pre-treatment BMI at the onset of cART, longer duration of HIV infection, low CD4 T cell count and white race (Leclercq et al., 2013). However, significant improvements in subcutaneous fat continued over 104 weeks after switching from a thymidine analogue zidovudine (ZDV) or stavudine (d4T) to abacavir (ABC) was observed after antiretroviral drugs intervention (Curran & Ribera, 2011).

Prominent facial lipoatrophy is the most well-known feature of lipodystrophy and becomes one of the affliction signs for PLWH on cART with good virologic and immunologic status. It becomes major reason for delaying the initiation of cART by patients and especially contributes to low compliance rate to therapy (Echavez, 2005).

Universiti Malaya

However, new scientific method using polylactic acid implants injection in PLWH patients with severe facial lipoatrophy proves successful in correcting the facial fat loss (Jagdeo et al., 2015). Polylactic acid is a biocompatible and immunologically inert synthetic polymer which belongs to the class of resorbable biomaterials.

Other factors associated with the onset of lipohypertrophy are time on antiretroviral therapy, time on cART regimens, change in viral load, prior treatment regimens, age, weight change, BMI and WHR (Falutz, 2011) and it may vary between women and men (Andany et al., 2011). Hence, it is assumed there are many factors influencing the syndrome and by varying degree.

With regards to ART drugs in Malaysia, NRTI and PI are no longer used as first line therapy anymore in clinics. There are now newer medications, combination tablets, once-a-day drugs and drugs with minimal long-term adverse events, in this case lipodystrophy syndrome. The newer generation of drugs are not associated with these changes and are much more acceptable to patients. For example, ART associated with lipodystrophy syndrome, d4T and other NRTI is switched to DTF or ABC, which may slow down or halt progression but may not fully reverse effects (Medical Development Division, 2011).

2.3.2 Metabolic syndrome

Metabolic syndrome is denoted by these criteria as defined by World Health Organization (WHO); fasting plasma glucose ≥ 1g/litre, fasting plasma triglycerides ≥ 1g/litre, large waist circumference (men: ≥ 102 cm, women: ≥ 88 cm), low fasting HDL-cholesterol (men: < 0.4 g/litre, women: < 0.5 g/litre) and high blood pressure (systolic > 130 mmHg and diastolic > 55 mmHg) (Saklayen, 2018).

Universiti Malaya

Metabolic syndrome development in PLWH is caused by multi risk factors which comprised of HIV-specific and antiretroviral-specific factors. Factor such as longer infection period, specific cART regimen and gender usually contribute to the development of metabolic syndrome, in addition to the traditional risk factors – obesity, tobacco use and genetic predisposition (Nguyen et al., 2016). So far, cART regimen becomes the focus on combating the metabolic irregularity in PLWH due to its mechanism even though its long-term effects are not fully understood. However, the new cART regimen is believed to pose fewer short-term metabolic agitation. The consequence of having metabolic syndrome is that it leads to increase in insulin resistance, dyslipidemia, hypertension and change in fat distribution. Moreover, combined with chronic inflammation and immune activation, it carries cardiovascular disease risk in PLWH. Present solution offered is cautious choosing of cART regimen where the regimen is both virologically effective and has minimum metabolic effects (Pontes Signorini et al., 2012).

2.3.3 Dyslipidemia

The cART therapy is linked to higher incident of dyslipidemia among PLWH compared to non-infected individuals even though the syndrome is lower during the early stages of HIV infection (Tripathi et al., 2013). Same as lipoatrophy and lipohypertrophy syndrome, the association of PI with dyslipidemia has been linked with decrease HDL-cholesterol level and increase triglyceride level (Grunfeld, 2010). Cardiovascular risk may also pose changes in lipid profiles in treating PLWH which leads to dyslipidemia but it depends on the type and duration of cART regime applied (Aberg, 2009).

However, one study involving Tanzania and Nigeria population proved otherwise. It is found that dyslipidemia is also possible without cART therapy (Adewole et al., 2010;

Universiti Malaya

Armstrong et al., 2011). Hypocholesterolemia (Míguez et al., 2010) and hypertriglyceridemia (Grunfeld, 2010) incidents are also reported with PLWH. Non-drug therapies have been suggested where dietary and exercise interventions have lowered the cholesterol level in significant amount (Husain & Ahmed, 2015).

2.3.4 Diabetes mellitus

Diabetes mellitus happens when intake of glucose into the body cell is interrupted and causes glucose for body cell use to decrease. This event is generated by lack of insulin sensitivity and decrease in insulin production where the former usually occurs in T2DM and the latter occurs in diabetes mellitus type 1. It is usually related to obesity syndrome caused by excess of fatty acid circulation which activate inflammatory response and concomitant chronic inflammation (Shi et al., 2006). In case of HIV infection, PLWH is reported to extract diabetes mellitus of 14% compared with 5% of non-infected individuals adjusted for age and BMI even though they are void of obesity (Brown et al., 2005). This phenomenon is suggested to originate from cumulative exposure to antiretroviral therapy with the strongest exposure caused from NRTI drugs. Moreover, length of viral exposure coupled with long period of cART regimens are two risk factors considered associated with diabetes mellitus in PLWH and it further rises with any insulin-glucose homeostasis disorder (Samaras, 2012). These risks are known to be associated with HIV comorbidities such as lipodystropy and dyslipidemia. Insulin resistance in PLWH is also associated with an increase in circulating free fatty acids where it contributes to intramyocellular lipid accumulation, an established risk factor for insulin resistance (Lake & Currier, 2013).

The PI class of antiretroviral, and in particular Indinavir and Ritonavir, may have a number of detrimental effects on metabolic health. Both have been shown to both increase

Universiti Malaya

insulin resistance through interference with glucose transport type 4 (GLUT-4)-mediated transport (Hadigan & Kattakuzhy, 2014). Ritonavir in particular has been shown to interfere with potassium signalling within β-cells impairing insulin release (Neye et al., 2006). Ritonavir in its most commonly used lower boosting dose is not associated with insulin resistance (Taylor et al., 2010). Indinavir and to a lesser extent ritonavir impair the ability of insulin to suppress endogenous hepatic gluconeogenesis (Lee et al., 2009). PI may also induce β-cell apoptosis. However, most PI developed more recently are believed not to have this effect (Hruz, 2011).

2.4 Body Composition in PLWH

Body composition can be divided to two major parts which are FM and fat free mass (FFM). First, FFM constitutes of bone mass and body lean mass and body lean mass is further divided into body cell mass and extracellular mass. Extracellular mass comprises of extracellular fluid and extracellular solids including the structural bone matrix (Earthman et al., 2000). In general, FFM is understood as all parts of body components but fat including body's water, bone, organs and muscle constituent. Whereas FM comprises of essential fat and non-essential fat, which is usually found in muscle tissue, under the skin known as subcutaneous fat or enveloping the organs known as visceral fat. The relationship between FM and FFM is found to be positive and linear after adjusting for body height, age and fat distribution (Burton, 2017).

There are several aberrations in body composition that occur resulting from the HIV infection. Continuous loss of body lean mass may also happen, and the consequence is a common occurrence of HIV wasting in PLWH (Tate et al., 2012). Patients who experience significant unintentional weight loss or weight loss that occurs at a rapid rate, irrespective of their BMI, required increased attention as they are at greater risk of

Universiti Malaya

morbidity and mortality. It is shown that FFM and body fat content of HIV-infected individuals were lower in comparison with normal individuals (Mutimura et al., 2010). However, there is no significant differences after age and height adjustment.

One study examines the body composition changes in men and women infected with HIV and it is found that FFM is dependent on the low initial percentage of body fat but FFM is not affected by the high initial percentage of body fat. It is different with women where the FFM is independent of the initial percentage of body fat regardless of any level of fat. (Forrester et al., 2002). But in these studies, no significant differences were found between PLWH and uninfected individuals in FM, body cell mass, BMI and intracellular water – extracellular water ratio (Mupere et al., 2010; Mutimura et al., 2010). In addition, women lose fat more than body lean mass during wasting and tend to preserve body lean mass until more advanced stages when significant loss of body cell mass also occurs.

As PLWH initiate their cART therapy, their body lean mass and body fat rise the first 96 weeks of therapy compared to non-infected individuals when body composition analysis was conducted using DEXA on these two groups. However, it changes after 96 weeks where PLWH continue to gain fat but their body lean mass decreases afterwards. (Grant et al., 2016) Testosterone treatment helps in addressing the loss of body lean mass and muscle mass even though further study needs to be conducted and on whether the combination of exercise and testosterone is more effective (Woerdeman & de Ronde, 2011).

2.5 Body Composition Assessment Method – Dual Energy X-ray Absorptiometry (DEXA)

DEXA has been widely used clinically to measure body composition in terms of FM and FFM. The is because it is safe due to low level of radiation exposure. The instrument

Universiti Malaya

generates two different X-ray energies which the differential attenuation of these two energies computes the soft tissue composition together with bone mineral content in the scanned area as shown in Figure 2.4. DEXA measurements are precise with 1% of coefficient of variation but it is still cannot be accepted as the "gold standard" for body composition. This is due to marked systematic differences found between different DEXA instrument systems (Bazzocchi et al., 2016).

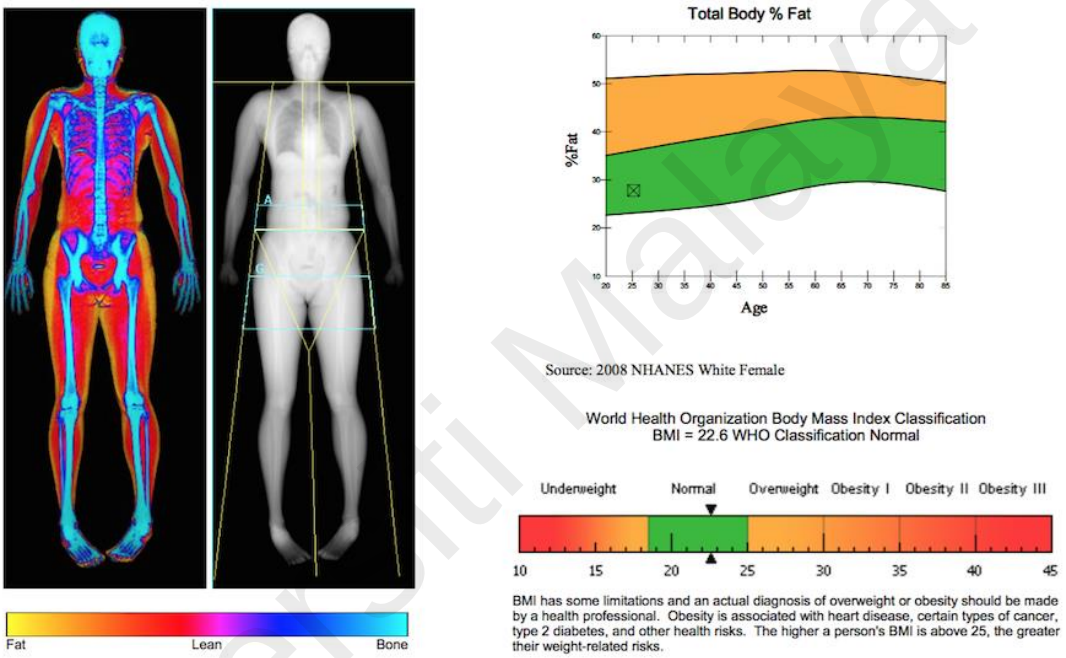

Figure 2-4 Whole-body DEXA scan of body fat, lean mass and bone (adapted from (Chor, 2015))

There are several assumptions on using DEXA to estimate FM. The first assumption is based on the following equation (Bazzocchi et al., 2016):

$$F_f = \frac{Rst - Rf}{R1 - Rf}$$

From the equation above, the determination of the ratio of the attenuation is described by attenuating pure fat (F_f) from FFM (R_1) relative to the soft tissue (R_{st}). In this case, R_f and R_1 are assumed as constant.

Universiti Malaya

The second assumption is that the measurements are independent from the anteroposterior thickness of the body. This assumption is considered as valid if the subject's thickness is < 20 cm. (Goodsitt, 1992) For subject with larger thickness, usually having obesity, it may violate the assumption (Jebb et al., 1993; Laskey et al., 1992). This is due to increased beam hardening which it is preferred for the lower energy X-rays to be attenuated (LaForgia et al., 2009). This causes the precision of DEXA body composition estimation to be reduced for broader subject over 100 kg (Milliken et al., 1996), in addition to difficulty in scanning due to the subject being wider than the scan area.

The third assumption is the edge detection differs throughout the body (Bazzocchi et al., 2016). Usually in clinical setting, edge detection by DEXA assumes that the soft tissue overlaying the bone as a constant composition. Moreover, each DEXA devices has different approaches to edge detection which may induces error between devices when measuring the soft tissue.

In order to overcome the challenges posed by the said limitations, half-body scans of each side of the obese individual's body was done and it is reported that the result agrees with the correlational analysis with hydrodensitometry for determining whole-body FFM, FM and body fat percentage (Brownbill & Ilich, 2005). But the body composition assessment is less accurate than whole-body assessment in healthy individual (Kiebzak et al., 2000).

Body composition evaluated by DEXA in untreated HIV-infected subjects show FM and body lean mass decreases even after adjustment on age, height, body bone mineral density. This finding is contributed by the low trunk FM while lower limb FM is high. (Delpierre et al., 2007) The result finding is similar with patients treated with cART, mainly PI and NRTI by measuring regional fat with DEXA in a longitudinal study (Mallal

Universiti Malaya

et al., 2000). However, it should be noted that the result may deviate in performing analysis due to issues of quality assurance (Smith et al., 2003) even though the reproducibility of DEXA as diagnostic tool is high regardless of populations when estimating body fat in HIV lipodystrophy (Cavalcanti et al., 2005).

2.6 Inflammation – Role of Interleukin-6 (IL-6) in PLWH

IL-6 is a proinflammatory cytokine designed to signal inflammatory markers such as C-reactive protein (CRP) (Wikby et al., 2006) or serum amyloid A (Thorn et al., 2004) during infection, acute injuries, autoimmune inflammatory diseases or malignancy. It is recognized as one of the glycoprotein chains with a molecular mass of 26 kDa (Tanaka et al., 2014). IL-6 is produced mainly by monocytes and macrophages, endothelial cells, fibroblast cells and plasma cells. It is responsible in stimulating hemostasis, influence T- and B-cell differentiation (Kerr et al., 2001), induce endothelial cells to produce chemokines (in complex with soluble IL-6 receptor (Sil-6r)) (Barnes et al., 2011) and assist in acute inflammatory response transition to chronic inflammatory response (Gabay, 2006) although it depends on age, smoking and acute phase response markers.

IL-6 is well-regulated and at low levels except in events cascading to inflammation including leukocytosis (Chen et al., 2011), thrombocytosis (Kaser et al., 2001) and acute phase protein synthesis (Gao et al., 2017). Cancer diagnosis and prognosis has integrated IL-6 as potential diagnostic marker especially in marking tumor progression (Vainer et al., 2018). Development of coronary heart disease, acute coronary syndrome and myocardial infarction has been linked with IL-6. Increase level of plasma IL-6 has become a risk factor and predictive marker to people with unstable angina pectoris according to these studies (Wainstein et al., 2017).

Universiti Malaya

Since IL-6 acts as inflammatory stimulator to the inflammatory reactant proteins, it is worth to note its relationship with FM. This is because FM components like elevated adipocytes, lower HDL cholesterol and elevated LDL cholesterol may increase IL-6 production (Gutierrez et al., 2009) and hence, the associated FM as a strong risk factor for inflammation. People with high FM level, for example obese people therefore tend to have higher plasma IL-6 level (El-Mikkawy et al., 2020). In addition, secretion of triglycerides is induced by IL-6 activity as shown by research in rats while high free fatty acid and triglyceride concentration are detected in human (Khovidhunkit et al., 2004). This observation is similar for LDL cholesterol levels which are associated with increased IL-6 activities (Robertson et al., 2017). Common metabolic disorders, such as diabetes, obesity, metabolic syndrome and hypertension displays abnormalities and peculiar trends in lipid and lipoprotein metabolism. The similar phenomenon is observed during infection and inflammation (Fuentes et al., 2013). It can be said the inflammation cytokines IL-6 contributes either directly or indirectly towards alteration in lipid and lipoprotein metabolism.

Elevated and sustained immune activation and inflammation are important characteristics of HIV infection (Paiardini & Müller-Trutwin, 2013). Increased expression of activation markers (CD38 and HLA-DR) on CD4 and CD8 cells as markers of immune activation have been observed in HIV infection (Appay & Sauce, 2008) in addition to increased expression of inflammatory cytokines IL-6, IL-4, and TNF-α in plasma and lymph nodes in the early stage of infection (Osuji et al., 2018). Localized inflammation in adipose tissue has been identified in adipose tissue changes and irregular behavior with multifactor is identified as the contributor. These alterations lead to erroneous adipokine release in endocrine function in adipose tissue, extra secretion of fatty acid and increase pro-inflammatory cytokines production such as IL-6, TNF-α or IL-Iβ (Xu et al., 2003).

Universiti Malaya

Findings show the level of pro-inflammatory cytokines in systemic body regulation is high in PLWH diagnosed with lipodystrophy (Singhania & Kotler, 2011) due to the action of cART stimulating the release of pro-inflammatory cytokines in adipocytes (Lagathu et al., 2007). This condition interrupts body composition as cytokine has reverse inhibitory effect on adiponectin, protein in adipose tissue and supply the lipodystrophy syndromes in PLWH (Lihn et al., 2003). This is further convinced by a research done on 2003 (Johnson et al., 2004). In addition, the inflammation itself either local or systematic causes extensive macrophage infiltration and develops enhanced production of pro-inflammatory cytokines (Bourgi et al., 2018) to combat and neutralize the site of inflammation and at the same time leads to change in adipose tissue composition. This phenomenon occurs in PLWH regardless in full- or naïve- antiretroviral therapy (Ross et al., 2009).

HIV-associated inflammation is also considered as a factor in increased level of chemokines involved in insulin regulation (Hruz, 2011) and in one study, it is found that higher levels of TNF-α markers were be significantly associated with the development of diabetes mellitus after first year on cART (Brown et al., 2010). Furthermore, increased microbial translocation across the gut in HIV patients further exacerbate inflammation and at the same time, correlate with the insulin resistance (Pedersen et al., 2013).

Elevated levels of pro-inflammatory cytokines, IL-6 is associated with atherosclerosis and subsequently, CVD such as myocardial infarction and stroke (Vos et al., 2016) as activation and progression of inflammation develops plaques formation. Ongoing inflammation despite undetectable HIV RNA levels also contributes to endothelial dysfunction in PLWH (Hsue et al., 2010) and further increases the CVD risks. Hence, increasing recognition of disturbed FM physiology is strongly associated with adipose tissue mass and implicated with IL-6 in PLWH. It is also found that systemic

Universiti Malaya

inflammation is greater in HIV-positive with lipodystrophy compared to insulin-resistant obesity patients (Samaras et al., 2009) which places PLWH as a group with high risk of diabetes.

2.7 Review on Past HIV Research Applying ANN as Their Methodology

Significant efforts have been dedicated to applying ANN methodology as a modelling technique for the complicated task of data analysis in HIV research and therapy in recent years. So far, the application of ANN in the field of HIV includes HIV biology and AIDS models, design and discovery of anti-HIV drugs (Bonet et al., 2007; Drăghici & Potter, 2003; Fogel et al., 2015) and pattern recognition in clinical diagnosis (Trikalinos et al., 2003).

ANN has been used to design anti-HIV vaccine drug by utilizing certain protein segment of the retroviruses to interrupt its working and infection mechanism. In this study, classification method is applied using bidirectional recurrent neural network. The 3D-structure of HIV-protein is used as the feature in the networks and it is found the use of segments of HIV-protein to predict HIV genotype had accuracy between 81.4% and 94.7%. The researcher propose that by classifying the HIV genotype, it become easier to predict the retrovirus resistance to ART (Bonet et al., 2007).

ANN has also actively been used in finding effective use of antiretroviral drug, PI to create more specific and less toxic PI (Drăghici & Potter, 2003). The study applied two methods to predict the HIV drug resistance using features in HIV protein to identifying mutation. The first method used cross-validation technique with structural feature of HIV protease and achieved accuracy between 60% and 70%, while the second method used ANN with sequence feature of HIV protease as its input. Several networks architecture developed by this second method were combined and it reached accuracy of 85%. The

Universiti Malaya

result shown that the mutation in HIV protease which led to drug resistance can be classified with high accuracy.

Whereas in another study, protein sequence of HIV was represented by 73 structural, biochemical and regional variables (Fogel et al., 2015). The variables were evaluated using feed-forward neural network with 15 input nodes, 3 hidden nodes and one output node in order to correctly classified the type of HIV protein that binds to immune CD4 cells. The hidden nodes used sigmoid activation function with initial sigma 0.1, initial weight 0.0 with inputs normalized to [0.1, 0.9]. The accuracy achieved by the ANN was 81.8% by using mean squared error (MSE) for measurement.

By modelling the disease progression using combination of clinical, metabolic and imaging/body composition data, ANN also improves the ability to detect potential risk factors and diagnose PLWH performance in identifying lipodystrophy (Trikalinos et al., 2003). Three-layers feed-forward backpropagation networks with different number of neurons in these layers were trained and validated which the database used for training were split into subject with lipodystrophy and subject without lipodystrophy. The networks were separated into three different type of models: network models with clinical variables only, network models with clinical and metabolic variables and network models with clinical, metabolic and body composition variables. The performance of network for the first two were similar at 71.8% accuracy and the performance of network with three kind of variables were higher at 78.5% accuracy.

Other application of ANN is to predict the retroviruses response to cART and prevent viral drug resistance to avoid treatment failure (Larder et al., 2007). Three-layers feed-forward ANN has been developed to predict the retrovirus response to ART using HIV genotype, baseline viral load, follow-up viral load, baseline CD4 cell and four treatment history variables with retrovirus response (current viral load) as output. Correlation

Universiti Malaya

between actual and predicted change in viral load starting from baseline and mean absolute difference between actual and predicted retrovirus response were used as the evaluation of the model performance. The network models achieved 0.69 correlation between the predicted and actual retrovirus response.

Another study utilized ANN to predict the ongoing immune CD4 cell count in PLWH using genome sequences, viral load and time period between current and baseline CD4 cell count in weeks as variables to track the progression of HIV infection (Singh et al., 2013). The network classified the output into either CD4 < 200 and CD4 ≥ 200. The study employed single layer recurrent ANN with eight neurons. The performance of the networks achieved accuracy of 95% which shown the ability of ANN to predict CD4 cell count in PLWH.

Three-layers ANN using backpropagation technique has also been developed to predict the incidence of HIV cases in China (Li & Li, 2020). The incidence data of HIV cases for every three years was used as input in training set to predict the incidence cases at the fourth year in validation set starting from year 2004 until 2016. The incidence cases of year 2017 then used as test (holdout) set to confirm the model performance. The model shows almost zero for minimum mean squared error value by comparing the predicted value of HIV cases with the actual incidence number. This outcome demonstrated the ability of ANN to model the actual incidence of HIV cases.

Application of ANN has been considered in identifying and grouping PLWH into groups with or without Hepatitis C virus infection (Rivero-Juárez et al., 2020). By identifying the co-infected HIV/ HCV patients, physicians are able to prioritize initiation of treatment with this group. Six characteristics variables were fed into the network in order for the model to classify the treatment group. The model used three layers of nodes with radial basis function which the weight was initialized in the range of [-5, 5]. The

Universiti Malaya

number of nodes and connections were set up within range [1, 2] including for hidden nodes. The overall performance was deemed good with 0.767 accuracy, 0.550 minimum sensitivity and 0.802 for area under ROC curve.

2.8 Non-Linear Modelling Method – Artificial Neural Network (ANN)

ANN, a non-parametric method functions analogously as human brain where its basic unit of a neural network is similarly mapped and imitated according to a human neuron network. It can be defined as an artificial interconnected neural network of neurons consisting of several processing layers that gain knowledge through analyzing and processing input information feed to them and translate it to a defined output by modelling human brain neural network behavior (Mhatre et al., 2017). Figure 2.5 shows the structure and functionality of a single artificial neuron.

The neuron system is fed a set of inputs that are symbolized by (X_1, X_2, X_3, ….. X_n) and a set of weights (w_1, w_2, w_3, ….. w_n) is multiplied with the set of inputs before the weighted inputs are summed and put into a non-linear function. Various type of activation functions can be applied, for example hyperbolic-tangent function, gaussian function, logistic function, arc tangent function and sine function. Input of 1 act as a bias and accounts for the factors that are not considered by the set of inputs. The output value, y is generated based on the output function from the layers previously which performs its computational mechanism by specified mathematical functions (Tanty & Desmukh, 2015).

Universiti Malaya

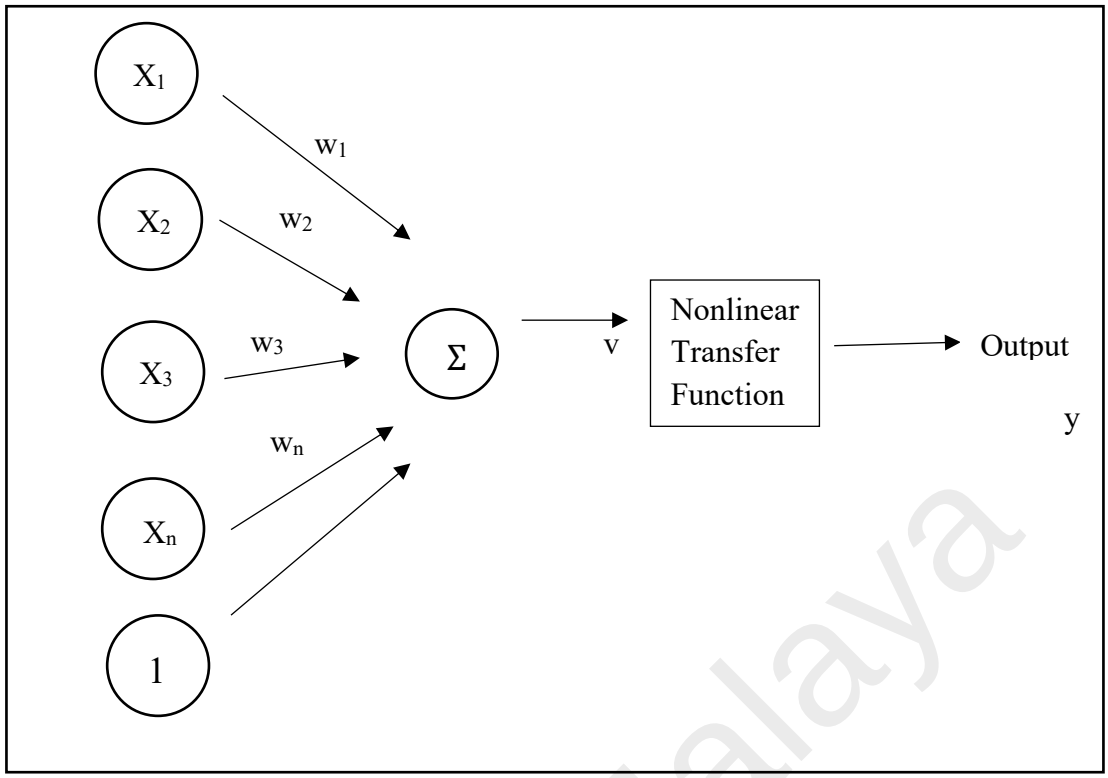

Figure 2-5 Structure of a simple artificial neural network model consisting of multiple inputs (X_1, X_2, X_3, …… X_n), correspomding weights (w_1, w_2, w_3, ….. w_n), one neuron, non-linear transfer function and single output y (adapted from (Tanty & Desmukh, 2015))

There are three types of processing layers which are input layer, hidden layer and output layer. The nodes in input layer usually have their values set as dependent variables which also serve as initial values for the network. The hidden layer connects the input layer and output layer where most of mathematical computing process occur in this layer since it supports the necessary function from input and output. The output layer computes the linear combination of activated weighted sum of nodes to generate required output and solve the complex problem it been given to (Sordo, 2002).

One of ANN application is feed-forward neural network. Deep learning model in supervised learning environment integrates feed-forward neural network which later known as multilayer perceptron as shown in Figure 2.6. The networks usually popular in area employing computer vision – image and video processing by computer by imitating

Universiti Malaya

human mind – or natural language processing – convert language to machine command – where neural network like convolutional neural network is applied to the design (Gupta, 2017).

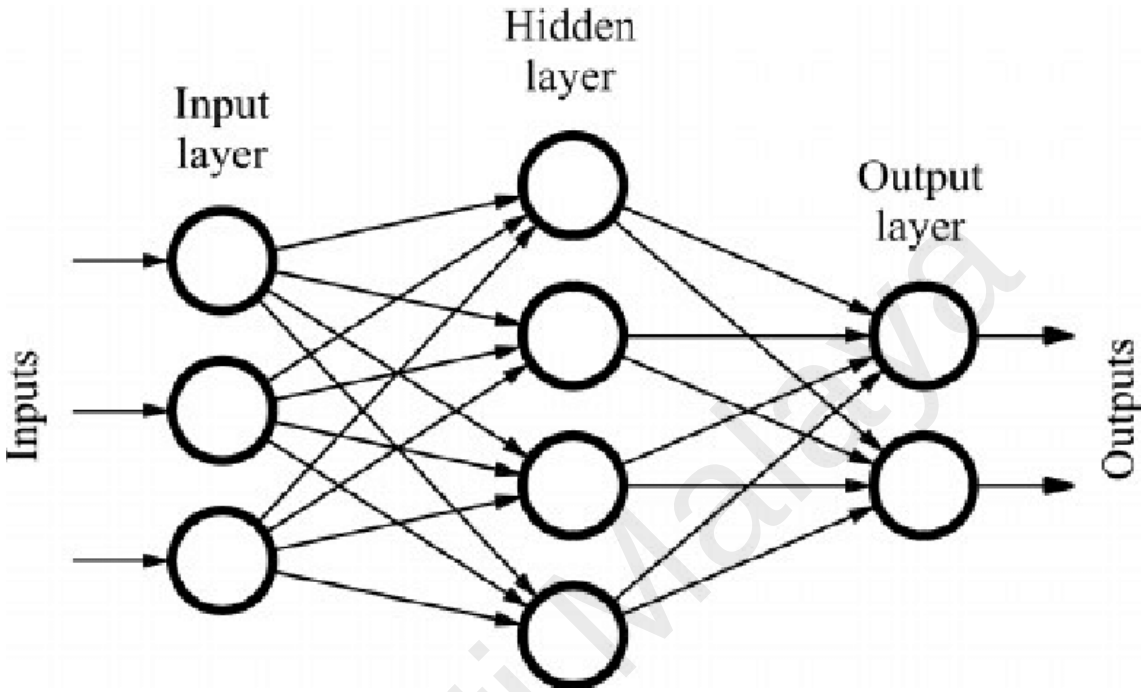

Figure 2-6 A simple feed-forward neural network with single hidden layer (adapted from (Sordo, 2002))

The feed-forward network is constructed to measure regression function, $y = f(x)$ while at the same time learns the best approximate value for function f without jeopardizing the non-linear function. The flow of information, x occurs in forward direction starting from input layer, hidden layer and output layer. Feed-forward neural network in this sense is different from recurrent neural network where the connection between its nodes form loop or loops with each other.

A back-propagation algorithm is designed and developed to train the feed-forward network (Li et al., 2012). It is used to counter the problem arises with how to derive the connection weights and tweak the values so that it can fit the intended result. The network output which is known as predicted output ($y_1 \ldots y_n$) is compared with the desired output

Universiti Malaya

($d_1 \ldots d_n$) and the difference between the two outputs is compared and noted as error of the network. The error is back propagated, and the network weights consequently change its value according to the present error which the same process happens to the other connection weights in every layer. This manipulation process of weight values and neuron values follow delta rule, a gradient-based weight adaptation method. The process is also known as an iterative process. The whole training process is repeated several cycles or iterations until the output error is reduced to an acceptable value. The trained neural network once configures the pattern of the problem during the training process will easily spot and identify the similar pattern in new problem and situation it come across and obtain the appropriate output which can be used for prediction or classification.

2.9 Comparison of Conventional Statistical Method and Artificial Neural Network

2.9.1 Introduction

Function of ANNs mainly can be divided into three. These functions are assigned according to the nature of the task carried out by said networks (Mhatre et al., 2017):

1) As models of the biological nervous system or in the analysis of intelligence in the area of neurophysiology.
2) As adaptive signal processing, image compression, speech recognition as well as automatic control application in the area of engineering or computer science.
3) As data analysis in the area of health science.

Therefore, ANNs were invented as a model mimicking the human thinking, tool for cognitive modelling and as a data analysis technique. These cognitive models surpass its counterpart, the statistical models via the complex architecture. The approach in which these networks are being trained is every so often part of the model and the assumptions

Universiti Malaya

that are being made. Thus, the comparison between ANNs and statistical models should be made based on ANNs as a kind of "data analysis" tool, which makes ANNs similar or identical to well-known statistical methods.

Between the two data analysis methods, difference and similarity exist and hence, need for comparison are addressed (Shahid et al., 2019). By using statistical method, researchers use human intelligence and not artificial intelligence in order to understand the process under analysis, identify the problems in the model and data, and investigate the outcomes. By contrast, ANNs acting as artificial intelligence render the networks to be "black box" which do not require human intervention to present the intrinsic learning and intelligence ability between data input and prediction output. On the other hand, the conventional statistical models stick with rigid assumptions about model structure that they are trying to assess from the database fed to them. ANNs, on the other hand are analogous to nonparametric methods (Jothi & Husain, 2015). The networks make no assumptions about the distribution of the data and no rules to be followed when assessing the problem at hand.

2.9.2 Model Formulation and Problem Solving

A mathematical expression for a linear model is as follows:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_n X_{in} + \epsilon_i = \sum_{j=0}^{n} \beta_j X_{ij} + \varepsilon_i \qquad \text{Equation 1}$$

where

Y_i : the value of dependent variable for the *i*th observation
X_{ij} : The value of the *j*th independent variable for the *i*th observation
β_j : The unknown parameter for the *j*th independent variable
ε_i : The residual term which is assumed to be E(ε) = 0, V(ϵ) is constant

Universiti Malaya

A linear model such as linear discriminant function or logistic regression is deriving from function relating variables X to output Y. By minimizing error measure or maximizing function, the coefficients attached to variables X are acquired to solve the problem (Stapleton, 2009). Whereas the ANNs resolves the trouble in term of heuristic search. The neural network starts the search for minimum error by setting the weights (equivalent to the coefficients in statistical model) randomly. An observation in the data set is then randomly chosen to present through the input layer to the output layer. The weights are then updated to decrease the error according to the difference between the desired output and the actual output. After all observations have presented in the network, the updated weights tend asymptotically towards describing the underlying relationship. The iterative procedure will find a minimum in the error function and yield result similar to those generated by conventional statistical methods (Gupta, 2017). However, there are some differences (Dreiseitl & Ohno-Machado, 2002):

1) Conventional statistical techniques necessitate some assumptions to derive the model. The method needs normality in the distribution of independent variables and the homogeneity of variance-covariance across groups. Whereas the ANNs algorithm does not depend on any specific distribution of the variables and the assumptions of statistical models. In other words, ANNs do not share the assumptions of normal distribution in independent variables and do not require linearity constraint.
2) In order to provide the overall output value, the ANNs may fit linear functions in the initial step through the hidden layer, which are then fed into the other function in the other hidden layer instead of fitting the desired relationship by means of just one unique function. In addition, ANNs approach inclines to be good at apprehending multivariate data and differentiate various outcomes, while the statistical models emphasis on capturing a single pattern and breakdown the

Universiti Malaya

explanatory variable estimate into parts which can be separately estimate, rather than on recognizing discriminating patterns.

The main benefit of a multi-layered ANNs rests in its capability to create the complex relationship representing the non-linear aspects in the problem (Landi et al., 2010). Meanwhile, statistical method necessitates that the decision set used to differentiate between groups must be linearly separable. It is likely to choose a cutoff point that is above or below this value for a single input and observation can be categorized into the correct group. However, when using two or more input variables, the inputs cannot be entirely differentiated by a single line or by a single plane because of the nature of multivariate inputs to form a nonlinear relationship. This makes statistical method unsuited with complex decision problem. On the other hand, multilayer perceptron model is a standard and flexible nonlinear model. It is a common approximation approach and can fit any function to any degree of accuracy if suitable hidden nodes are chosen even with unknown relationship between the independent and dependent variables.

2.9.3 Statistical Testing and Interpretation

The significance testing of the model and the evaluation of independent variables are important parts of statistical methods. The multilayer ANNs with suitable hidden units can accurately model the relationship between input and output variables. However, it will lead to overfitting and that will affect the solution for problem given for training (Ying, 2019).

Other criteria need for consideration is the adaptability of model. A model is required for its underlying relationships and parameters to be stable over time. Given ANNs are able to adapt to the alteration in the predictive model when new observations fed into the model, it is preferable to employ ANNs in prediction problem (Sarvepalli, 2015). When

Universiti Malaya

the current distributions are changing, ANNs still use the past data as the data is gradually reduced in importance when new data are fed into the network. This adaptive learning process is well-suited with real world and believed to be valuable part of its effectiveness. Whereas the statistical methods need batch update and the old data is employed to construct a new model when new observations are entered (Dreiseitl & Ohno-Machado, 2002). Hence, statistical methods cannot adapt itself to the feature of gradual transformation of the environment since statistical models presume that old and new data are equally important.

2.9.4 Performance Comparison Between Conventional Statistical Method and Artificial Neural Network

A brief overview of performance comparison of ANN with conventional statistical methods were presented in this sub-chapter. Based on the considerations mentioned in previous sub-chapters (Model Formulation and Problem Solving and Statistical Testing and Interpretation), it seems appropriate to attempt a formal comparison of the two approaches (statistical method vs. ANN).

ANN was applied to model the mortality and morbidity in PLWH during loss of HIV T-Cell homeostasis (Hatzakis & Tsoukas, 2002). Feed-forward backpropagation ANN was developed to assess the outcomes of therapy using multiple risk factors, including initial response to therapy, viral factors and host immune parameters. The network is a single hidden layer with one input and output layer with sigmoid transfer function. The performance of ANN was validated using Cox regression modelling where the comparison result showed both methods are successful in predicting the desired outcome even though ANN performed better in assessing the risk that occurs in late stages of HIV infection due to loss in T-Cell homeostasis.

Universiti Malaya

Prediction analysis of quantitative structure-activity relationship was conducted using ANN where the study used inhibition constant of 127 symmetrical and unsymmetrical cyclic urea and cylic cyanoguanidine derivatives containing different substituent groups (Deeb & Jawabreh, 2011). Then, multiple linear regression analysis with stepwise selection and elimination of variables was used to model the relationship of HIV viral load with the rest of descriptor groups. The study concluded that the models from ANN analysis performed better than the multiple linear regression models by using external and cross-validation technique to validate the performance of both methods.

Another study applied ANN to improve accuracy to predict time series data in HIV (Eswaran & Logeswaran, 2011). The ANN model was compared with statistical model, moving average and auto regressive integrated moving average model. Root mean square error and mean absolute error were used as measures to compare the performance of both models. The result showed that ANN model performed better compared to other statistical models. Significant improvement was present and ANN model was proposed a better model in HIV time series prediction.

Comparison of several prediction techniques was explored in development of HIV testing prediction model by using ANN, decision tree, naïve bayes and logistic regression technique (Hailu, 2015). The result indicated that ANN model performed better compared to regression model even though the performance of ANN was still small from decision tree models which reported as having best performance.

Almost the same analysis was carried out in 2019 to explore the occurrence of HIV in time series models (Wang et al., 2019). Long short-term memory neural network, autoregressive integrated moving average (ARIMA), generalized regression neural network and exponential smoothing were used to model the incidence data. The model performance was evaluated using mean square error, root mean square error, mean

Universiti Malaya

absolute error and mean absolute percentage error. Prediction performance of long short-term memory neural network models were reported good as compared with other models. The author relates the good predictive performance to the model capacity to model the long-term time series data and determine the optimal time lag when the time lag is long and have unknown size.

Prediction of 6-year incidence of metabolic syndrome was attempted by using three types of ANN: multilayer perceptron, radial basis function and linear networks and multiple logistic regression (Hirose et al., 2011). The analysis used clinical risk factors, including the insulin resistance index calculated by homeostasis model assessment. It was found that performance sensitivity of ANN model was higher compared to multiple linear regression model. The analysis also identified BMI, age, HDL-cholesterol, LDL-cholesterol, diastolic blood pressure and insulin resistance parameters as important predictive risk factors. It is concluded that ANN is a good method to predict the 6-year incidence of metabolic syndrome based on clinical data.

2.10 Summary

The use of ART has resulted in greater improvements in morbidity and mortality in PLWH. However, the use of ART regimen has its own consequence with reports of metabolic abnormalities such as lipodystrophy syndrome, metabolic syndrome, dyslipidemia and insulin resistance. Alarming morphologic changes in PLWH associated with these metabolic abnormalities usually characterized by accumulation of fat in the abdomen and in the dorsocervical area of the neck, as well as by the depletion of fat in the face, buttocks and extremities. The reason associated with the metabolic disturbances and morphologic changes related to RT are not understood completely. The etiology is likely to involve side-effects of some ART drugs and the inflammatory effects either

Universiti Malaya

caused by the regimen or the HIV infection itself. This is so when increased expression markers of immune activation have been observed in HIV infection taking ART.

In medicine, ANNs have been used extensively in medical diagnosis, detection and evaluation of medical conditions. Furthermore, ANNs have also been applied in data mining projects for the purposes of prediction, classification, and response modelling. Traditionally, statistical techniques are employed in the analysis before ANNs are introduced into the field. The predominant technique is logistic regression in which the effects of predictors are linear. It is difficult to use these models to discover unanticipated complex relationship, i.e., non-linearities in the effect of a predictor or interactions between predictors. ANNs have the potential to discover non-linear relationship which can be used in data analysis problem.

Universiti Malaya

CHAPTER 3: METHODOLOGY

3.1 Introduction

In this chapter, the research methodology used is outlined. The data collection and data completion together with subjects are described in this chapter. Data analysis using non-linear modelling techniques was constructed using ANN and its model design and development are also described in this chapter.

3.2 Data Collection

The study was conducted using data acquired through the Malaysian HIV and Aging Study (MHIVA) in collaboration with the Biomedical Engineering Department, Engineering Faculty, University Malaya. Details of the study have previously been published (Rajasuriar et al., 2017). Briefly, the study encompassed recruitment of PLWH who were on routine follow up at the Infectious Diseases Unit, University Malaya Medical Centre. All participants fulfilled the following inclusion criteria; age > 25 years, on suppressive ART for at least 12 months and had no acute illness at recruitment. Participants consenting to the study had detailed biochemical screening including assessment of fasting lipids and glucose and anthropometric assessment while relevant HIV- related parameters were extracted from each participant medical records. A subset of participants also had whole-body DEXA performed. Due to logistics constrains, separate appointments were provided for imaging analysis and all participants had DEXA scans performed within 6 months of recruitment. All participants provided informed consent and the protocol for the study was approved by the institutional review board (MEC 20151-937). Bloods were also collected in EDTA vacutainers and processed within 4 hours of collection to isolate plasma as previously described (Yap et al., 2017). Levels of IL-6 were measured by cytokine bead array (BD Bioscience, USA) according to the manufacturers' instructions.

Universiti Malaya

The data generated from this study was provided in the form of .xlxs files which was converted to .sav files for analysis in SPSS. First part of the data consists of clinical / demographic parameters, metabolic parameters and inflammation marker parameter while the second part of the database was obtained from iPacs system database of University Malaya Medical Centre which stores the body composition parameter measured using DEXA. The first part of data contains of 327 cases and the second part of data contains 115 cases where both parts of data were cross matched with each other and selected for analysis.

3.3 Data Pre-processing

The cases in the first part of data were cross matched with the cases in the second part of data according to the case number and the number of cases obtained were 107. However, the data obtained for this study contained incomplete data due to the metal error (11 cases) and missing data (21 cases). The data and its case number were identified and eliminated which brought the number of cases to 86. Then, female patient cases were excluded from the analysis because of small sample size (15 cases) which could affect the analysis because of the gender skew. The final number of complete cases were 71 cases after cross-matching and elimination as shown in Figure 3.1.

Data is prepared and cleaned by removing outliers, missing data, irrelevant data and inconsistent data. Formatting of the data was also conducted so that variations in data did not treated as different classifications during processing. The inconsistent data was normalized to ensure the input feature data was on the same scale and comparable. It is to prevent bad performance by the neural network if different attributes were on different scales which can caused some of the attributes to be counted more than the others and hence, caused bias. This technique was conducted on baseline and current HIV viral load

Universiti Malaya

parameters where the values were converted to logarithm with base 10. ANN analysis by SPSS software also came with its own pre-processing modules with normalize and scale functions.

For the data fed into the network were of quality data, outliers were removed during data cleaning. Outliers were detected by computing the standard deviation of the intended data and making sure the data distribution was normal. A data point was considered as outlier if the data value were outside the standard value or it affected the distribution if it is included with other data points.

Approach taken for error tolerance is by expanding decimal points of data values to four decimal places and other approach taken to decrease the error on user side was by preparing and cleaning the data before it is fed into the network for training as explained in previous paragraphs. Variance inflation factor test was also conducted to measure the multicollinearity among the variables.

Universiti Malaya

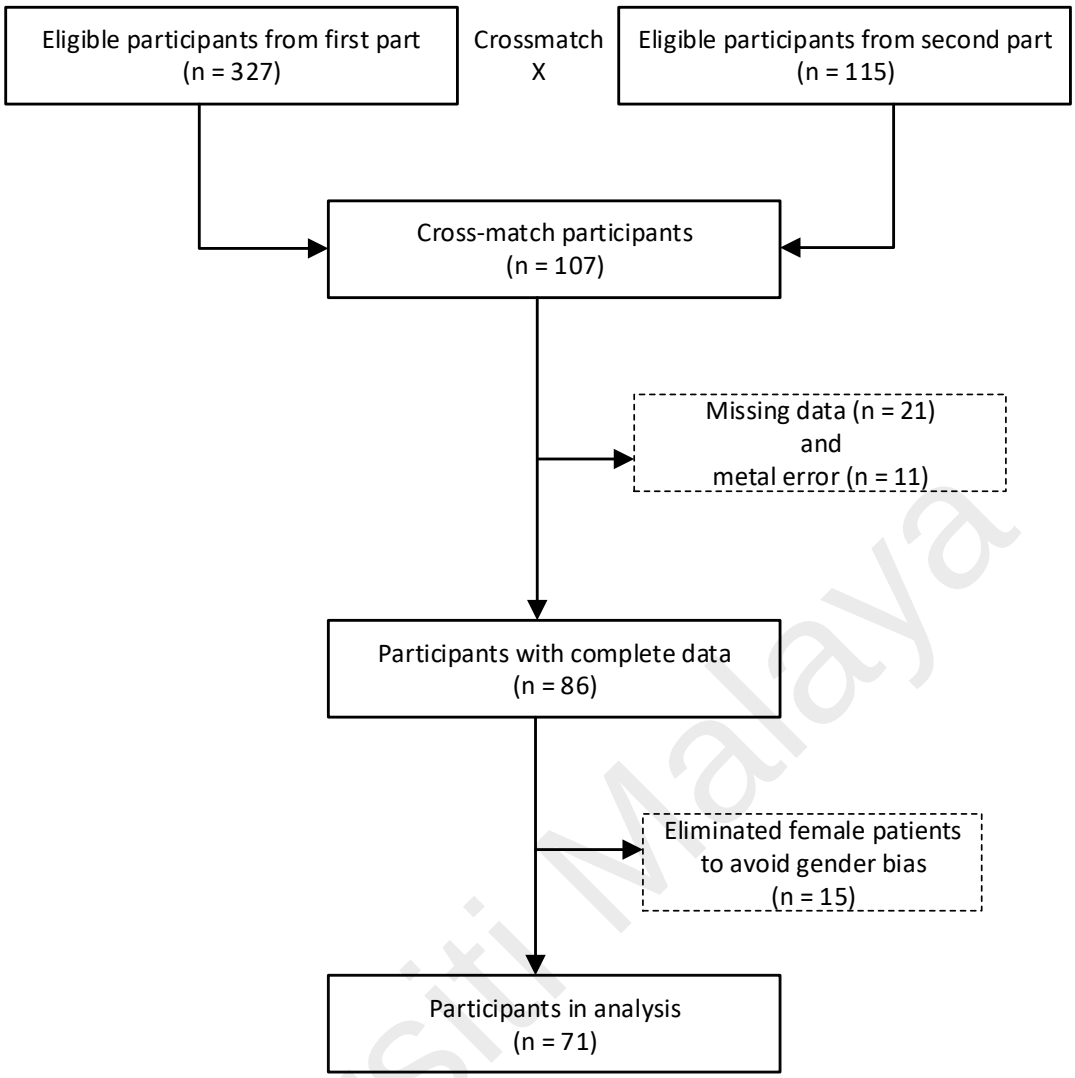

Figure 3-1 Flow chart of data cleaning process for the analysis

3.4 Data Analysis

Descriptive statistics were calculated for the variables involved in this study comprised of FM, inflammation marker (IL-6), clinical and metabolic variables and are expressed as means ± standard deviation (SD). Bivariate linear regression analyses were performed to test simple hypotheses of association between FM and clinical, metabolic and inflammation variables. All tests were two-tailed and a p value < .05 was considered as statistically significant.

ANN modelling of relationship between body fat composition and inflammation involves the interaction of many diverse in nature variables. However, this study only

Universiti Malaya

utilized several variables to be used as independent variables. This action was taken to prevent ill-defined and ill-structured complex relationship between variables potentially able to influence the dependent outcome. For this purpose, fifteen fixed variables were selected as independent variables. These independent variables consisted of clinical variables (age, height, weight, WHR, BMI, administered duration of cART by months, baseline CD4 cell count, current CD4 cell count, $\log_{10}$ baseline HIV viral load and $\log_{10}$ current HIV viral load) and metabolic variables (fasting glucose, triglycerides, HDL, LDL and total cholesterol). The dependent variable consisted of body composition variable, FM collected from whole-body DEXA scan. Five sets of ANN models were constructed, and each set of network model was produced and manipulated based on inclusion of inflammation marker, IL-6 as independent variables and on statistical significant association of independent variables with dependent variable, FM. Comparison of model performance were assessed through the model error function, error sum of squares (SSE), relative error (RE) and mean predictive accuracy percentage (MPA%). The SSE was given by the following equation:

$$E = \sum_{n} \varepsilon_n = \sum_{n} (d_n - y_n)^2 \qquad \text{Equation 2}$$

where $\in_n$ is output error, d_n is desired output and y_n is predicted output.

MPA% was calculated using the following equation:

$$MPA\% = \frac{1}{N} \sum_{i=1}^{N} \left(1 - \frac{|FM_P - FM_D|}{FM_D}\right) 100\% \qquad \text{Equation 3}$$

where N was the number of subjects, FM_P was the predicted fat mass value estimated by ANN, FM_D was desired fat mass measured by DEXA method. Independent sample *t*-test was conducted on Set A and Set B to test the difference between the two ANN model sets. Another two statistical test were conducted to find the significant differences among

Universiti Malaya

ANN models in Set A, B, C, D and E which were ANOVA test and post hoc analysis using Tukey HSD test. Data was analyzed using software SPSS version 25 (IBM Corporation, Armonk, NY, USA).

3.5 Model Design and Building

Feed-forward back-propagation algorithm was developed for the ANN model involving three-layers topology acting as model processor. Each layer consisted of a perceptron produced when several of these neurons interacted with each other in a layer of linear threshold unit. Multilayer perceptron neural network was formed because these several layers of perceptron connected in complex interaction involving reinforcement of weights that led to correct behavior of model during training. The input layer received information which served as input to the neurons in the layer and passed the information to the neural network system for processing. The hidden layer received the information passed by input layer and performed deep learning when training the data while the trained information was sent to the output layer for final step. The bias neutron acts as fixed constant in the model training.

Gradient descent was applied to train the neural network model using chain rule. It measured the error of the model learning system by plotting a certain function that showed the possible results an ANN model can produce by different set of parameters. For each training step, the output error was computed, and the contribution of the error was calculated for each neuron in every layer. The current weight was used to backpropagate the computed error to individual connection of neurons and the information was used to tweak the weights through gradient descent (Dkhichi & Oukarfi, 2014). New value was then generated at the next pass or epoch of training with new optimized weight and

Universiti Malaya

connections of neurons. The training steps were repeated until the system converged and produced output with the least amount of error.

For training to take place, the data was split into three segments of data set – training data set, testing data set and holdout data set. The model was trained using only the data in training data set and learned the correct input-output response behavior to yield the desired output solution. The model was subjected to the test data set after the completion of model training to let the model tuned and tweaked itself with the varied distribution of input seen in training data set. The adjustment activity purpose was to make the model more reliable and robust. The performance of the model was evaluated using the holdout data set which was the separate untrained data to measure the model accuracy in predicting the correct output solution.

To further strengthened the ability of the model to perform the desired operation accurately, K-fold cross validation technique was applied to combat the overfitting problem (Koehrsen, 2018). Three segments of the K-randomly assigned data set segments were reserved as test data set and holdout data set. Each of remaining K-1 segments was individually trained and its performance was measured against the test and holdout data set. The average resulting error scores was taken as a final error metric from the K-fold cross validation.

Numbers of single-hidden layer ANN models were constructed with each of the models had different number of nodes in the hidden layer to select the optimal solution for the network system. Example of ANN diagram with four hidden nodes in its hidden layer was shown in Figure 3.2. No definite and exact protocols to be followed when choosing the optimal ANN architecture as long as the basic artificial neural network topology was followed (Karsoliya, 2012). Trial-and-error method was used in this study to determine the number of hidden nodes for the hidden layer (Sheela & Deepa, 2013).

Universiti Malaya

Hyperbolic tangent activation function was used in the hidden and output layer. The neural network function was used because it had wider range of real-value transformation and increased the possible derivation of output value letting faster training for the system convergence.

The training data was randomly divided into three partitions with ration of 7:2:1 for Set A, B, C and D. The three partitions corresponded to training dataset, testing dataset and holdout dataset. 7:2:1 split ratio was chosen because more data is needed for training, with less training data, the parameter estimates have greater variance. The testing data is also larger than holdout in order to reduce variance. However, to address this issue, one set of ANN models (Set E) were added and trained using 60% training and 40% testing ratio. The input used were similar variables with Set D. For Set E, training data was randomly divided into 2 partitions, 60% training dataset and 40% testing dataset. Many strategies have been adopted to stratify the division ratio for training, testing and holdout.

The common strategy is to split the data into 7:2:1 ratio as implemented by these studies (Gaj et al., 2020; Kornblith et al., 2020; Vang et al., 2018). These studies also employed cross-validation technique with the stratification strategy to avoid overfitting because of their small number of samples. The sizes of the partitions were different for every neural network model and it was self-assigned by the network algorithm. Network configuration required a set of initializing parameters to be set at random to initiate the training and certain value was assigned to the parameters as default to every ANN model. The value assigned to the learning rate was 0.4. The value was default value set by the system based on the small sizes of the sample and smaller value is used between range 0.0 and 1.0. Algorithm flowchart was presented in Figure 3.3

The output is represented as the following (refer Figure 2.5):

Universiti Malaya

$$v = \sum_n x_n w_n \quad \text{Equation 4}$$

$$y = f(v) \quad \text{Equation 5}$$

$$y = f(\sum_n x_n w_n) \quad \text{Equation 6}$$

The following derivation is adapted from (Baughman & Liu, 2014), improved and revised specific to this study. The hyperbolic tangent function of the output is:

$$f(x_n) = \tanh(x_n) = \frac{e^{x_n} - e^{-x_n}}{e^{x_n} + e^{-x_n}} = y \qquad -1 < y < 1 \quad \text{Equation 7}$$

Therefore, the partial derivative of the hyperbolic tangent function is:

$$\frac{\partial f}{\partial x_n} = \frac{(e^{x_n} + e^{-x_n})(e^{x_n} + e^{-x_n}) - (e^{x_n} - e^{-x_n})(e^{x_n} - e^{-x_n})}{(e^{x_n} + e^{-x_n})^2} \quad \text{Equation 8}$$

The equation can be rewrite to:

$$\frac{\partial f}{\partial x_n} = 1 - \frac{(e^{x_n} - e^{-x_n})^2}{(e^{x_n} + e^{-x_n})^2} \quad \text{Equation 9}$$

And therefore,

$$\frac{\partial f}{\partial x_n} = 1 - tanh^2(x_n) \quad \text{Equation 10}$$

The error is backpropagate through the network starting at the output and moving backward towards the input according to the equation:

$$\varepsilon_{n=}(1 - tanh^2(x_n))(d_n - y_n) \quad \text{Equation 11}$$

Universiti Malaya

Weight factor is adjusted, and new weight is introduced:

$$w_{n,new} = w_n + \eta i \varepsilon_n \qquad \text{Equation 12}$$

or

$$w_{n,new} = w_n + \eta i (1 - tanh^2(x_n))(d_n - y_n) \qquad \text{Equation 13}$$

The term η is a positive constant controlling the learning rate which in this study its value is 0.4. The learning rate value is randomly assigned by the network algorithm. Thus, the new weight factors are calculated from the old weight factors from the previous training iteration by the following general expression:

$$\begin{bmatrix} new\ weight \\ factor \end{bmatrix} = \begin{bmatrix} old\ weight \\ factor \end{bmatrix} + \begin{bmatrix} learning \\ rate \end{bmatrix} \times \begin{bmatrix} input \\ term \end{bmatrix} \times \begin{bmatrix} gradient - descent \\ correction\ term \end{bmatrix}$$

The equation is applied and repeated until the error sum of squares, *E* or output-error, ε is zero or sufficiently small.

Universiti Malaya

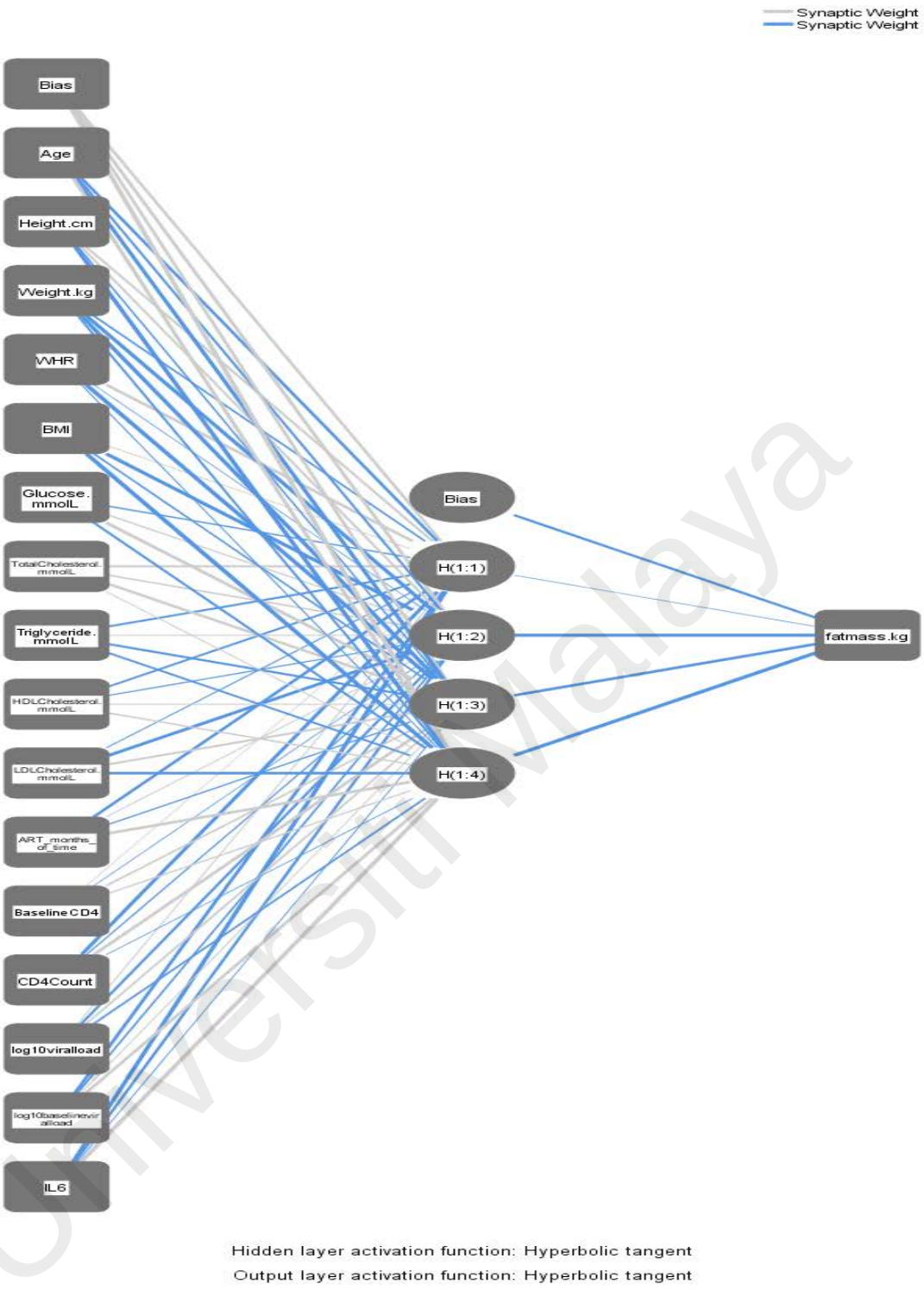

Figure 3-2 Example of ANN diagram with four hidden nodes in its hidden layer

Universiti Malaya

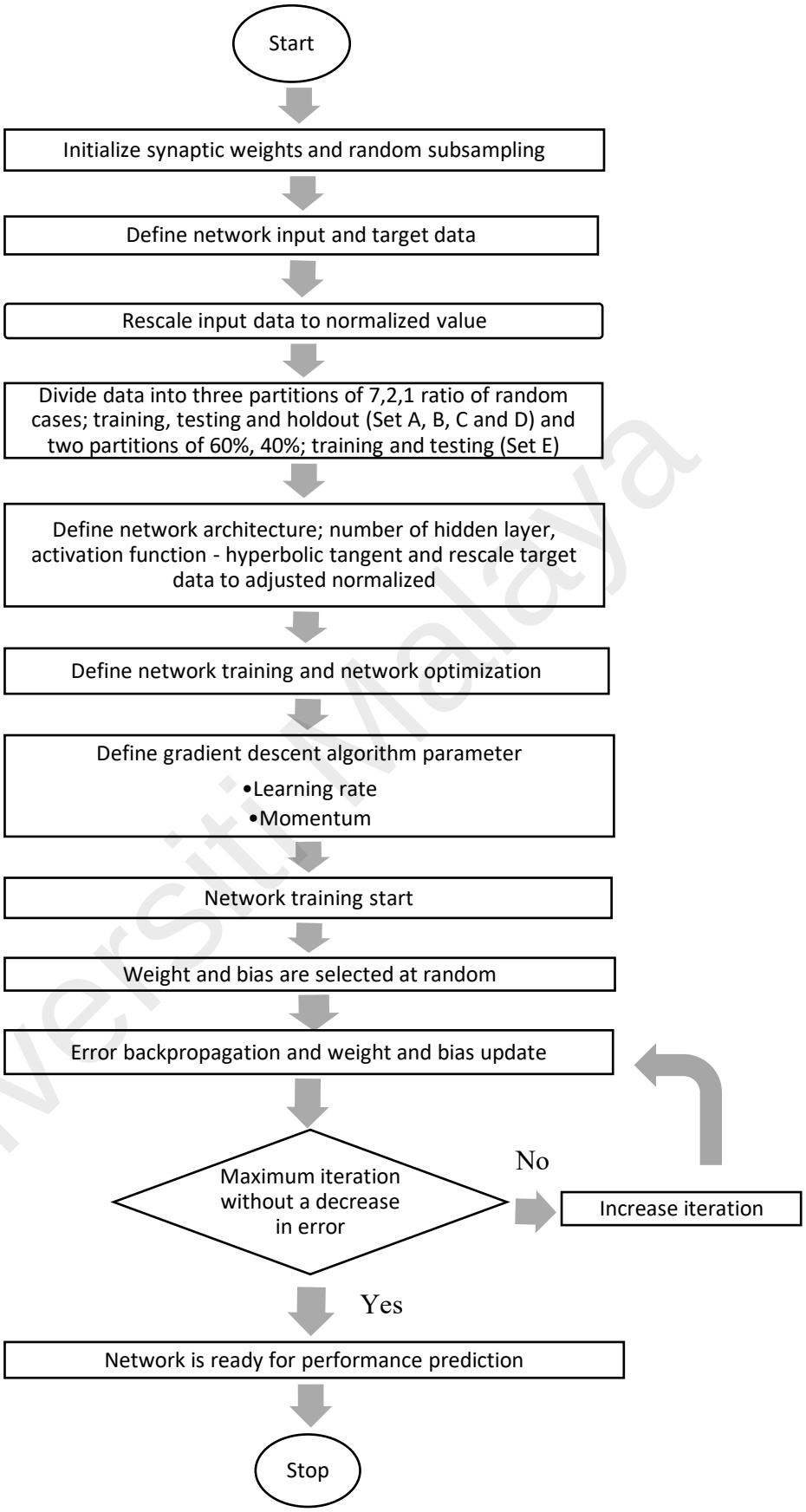

Figure 3-3 Flowchart of the multi-layer perceptron ANN

University of Malaya

3.6 Summary

The study was conducted using data obtained from the Malaysian HIV and Aging Study (MHIVA) in collaboration with the Biomedical Engineering Department, Engineering Faculty, University Malaya. The data consisted of clinical / demographic variables, metabolic variables, inflammation variable and body fat composition variable. Data preparation and cleaning were performed by removing outliers, missing data, irrelevant data and inconsistent data. The final number of complete cases were 71 cases of MLWH after cross-matching and elimination.

Descriptive statistics were calculated for the variables involved in this study comprised of FM, inflammation marker (IL-6), clinical and metabolic parameters and are expressed as means ± standard deviation (SD). Bivariate linear regression analyses were performed to test simple hypotheses of association between FM and clinical, metabolic and inflammation parameters.

Feed-forward back-propagation algorithm was developed for the ANN model involving three-layers topology acting as model processor. For training to take place, the data was split into three segments of data set – training data set, testing data set and holdout data set. Numbers of single-hidden layer ANN models were constructed with each of the models had different number of nodes in the hidden layer to select the optimal solution for the network system. Every set of ANN model were compared using statistical method to learn the significant difference of prediction accuracy among them.

Universiti Malaya

CHAPTER 4: RESULTS

4.1 Introduction

In this chapter, subject characteristics comprising clinical, metabolic and body composition for MLWH were described using mean and standard deviation. Following that, the analysis for bivariate linear regression were reported with FM as dependent variable while the rest of study variables as independent variable. After that, results from non-linear modelling analysis using ANN were reported where comparison and similarity among ANN sets were presented and described.

4.2 Descriptive Result

Table 4.1 shows the characteristics of MLWH patients in this study (n=71) comprising 17 variables. Four characteristics are selected where the first characteristic, clinical characteristic has ten components (age, height, weight, WHR, BMI, consumption period of antiretroviral medication, baseline and current CD4 cell count and baseline and current HIV viral load count), the second characteristic, metabolic characteristic has five components (glucose, triglycerides, total cholesterol, HDL-cholesterol and LDL-cholesterol), the third characteristic, body composition characteristic FM and fourth characteristic, inflammation characteristic IL-6. The dependent variables are clinical, metabolic and inflammation parameters whereas the dependent variables are FM parameter.

Universiti Malaya

Table 4-1 Subject characteristics comprising clinical, metabolic and body composition for HIV-positive male subjects. Dependent variables: clinical, metabolic and inflammation parameters. Independent variable: fat mass parameter. BMI-body mass index, WHR-waist-hip ratio, cART-combination of ART, CD4-cluster of differentiation 4, HDL-high-density lipoprotein, LDL-low-density lipoprotein, IL-6-interleukin-6.

Subject characteristics (n=71)	Mean ± SD
Clinical	
Age (years)	44.05 ± 1.22
Height (cm)	170.2 ± 0.80
Weight (cm)	66.70 ± 1.45
WHR	0.90 ± 0.01
BMI	23.07 ± 0.46
Period cART administered (months)	73.85 ± 6.21
Baseline CD4 count (cells/mm^3)	150.95 ± 15.52
Current CD4 count (cells/mm^3)	543.1 ± 28.73
Log_{10} baseline HIV viral load (copies/ml)	4.50 ± 1.52
Log_{10} current HIV viral load (copies/ml)	1.32 ± 0.10
Body composition	
Fat mass (kg)	17.28 ± 0.89
Metabolic	
Glucose (mmol/L)	5.66 ± 0.19
Triglycerides (mmol/L)	1.79 ± 0.12
Total cholesterol (mmol/L)	5.08 ± 0.11
HDL cholesterol (mmol/L)	1.33 ± 0.04
LDL cholesterol (mmol/L)	2.95 ± 0.09
Inflammation marker	
IL-6 (ng/mL)	2786.52 ± 303.31

For the clinical characteristic, the average patient age is 44.05 years old with 1.22 years deviation, the average height is 170.20 cm with 0.8 cm deviation, the average BMI is normal at 23.07 with 0.46 deviation. The average duration of treatment with cART is six years with six months deviation. The average CD4 count increases from 150.95 cells/mm^2 with 15.52 cells/mm^2 deviation from baseline to 543.10 cells/mm^2 with 28.73 cells/mm^2 at current average CD4 cells count. The average HIV viral load copies count where its count is measured in logarithm 10 is 4.50 with 1.20 deviation at baseline and drop to 1.32 with 0.10 deviation for current HIV viral load copies.

For the metabolic characteristic, the average fasting glucose value is 5.66 mmol/L with 0.19 mmol/L deviation which is borderline normal. The average triglycerides value is

Universiti Malaya

1.79 mmol/L with 0.12 mmol/L deviation which is in intermediate range, the average total cholesterol is normal at 5.08 mmol/L with 0.11 mmol/L while the average HDL-cholesterol and average LDL-cholesterol is low at 1.33 mmol/L with deviation 0.04 mmol/L and 2.95 mmol/L with deviation 0.09 mmol/L. The average body composition or FM is 17.28 kg with 0.89 kg deviation while the average value of inflammation marker, IL-6 is 2786.52 ng/ml with 303.31 ng/ml deviation.

Universiti Malaya

4.3 Bivariate Linear Regression Analysis

Table 4.2 displays the bivariate linear regression between FM and study variables comprising of clinical variables, metabolic variables and inflammation marker, IL-6 variable. Of note, moderate positive correlation is observed between FM value and WHR (r = 0.52). Additionally, weight (r = 0.91) and BMI (r = 0.85) show strong positive correlation with FM as predicted. Whereas the rest of variables are weakly positively correlated with FM as displayed in Table 4.2.

The table also provides the total variance, r^2 for FM that is explained by the study variables where weight and BMI explain large variation at 82.4% and 71.8% while WHR explain low variation at 27.1%. Meanwhile, there are no, or only very small variations associated between FM with the rest of the variables – age (0%), height (8.5%), duration on cART (2.3%), baseline CD4 (0.1%), current CD4 (0.3%), baseline HIV viral load (0.4%), current HIV viral load (6%), glucose (4.8%), triglycerides (7%), total cholesterol (0.3%), HDL (10.5%), LDL (0.3%) and IL-6 (0.2%).

In bivariate regression analyses, significant association is observed between FM and height (coef: 0.325 , SE: 0.129 , p = 0.014), weight (coef: 0.563 , SE: 0.031, p = 0.000), WHR (coef: 54.776, SE: 10.816, p = 0.000), BMI (coef: 1.645, SE: 0.124, p = 0.000), current HIV viral load (coef: 18.366, SE: 8.776, p = 0.040), triglycerides (coef: 1.974, SE: 0.864, p = 0.025) and HDL (coef: -7.309, SE: 2.565, p = 0.006). The linear regression models developed in this study did not demonstrate notable collinearity as indicated by variance inflation factor where the value is at 1.0 for all the study variables.

Universiti Malaya

Table 4-2 Bivariate linear regression analysis result for relationship between fat mass and each variable comprising of clinical, metabolic and inflammation marker, IL-6 variables in HIV-positive males (n=71)

DV	IV	r	r^2	*p*	coef.	SE	constant	VIF
Fat mass	Age	0.02	0.000	0.990	-0.001	0.089	17.350	1.0
	Height	0.29	0.085	0.014*	0.325	0.129	-38.077	1.0
	Weight	0.91	0.824	0.000*	0.563	0.031	-20.396	1.0
	WHR	0.52	0.271	0.000*	54.776	10.816	-31.999	1.0
	BMI	0.85	0.718	0.000*	1.645	0.124	-20.662	1.0
	ART duration	0.15	0.023	0.206	-0.022	0.017	18.942	1.0
	Baseline CD4	0.03	0.001	0.781	0.002	0.007	16.984	1.0
	Current CD4	0.06	0.003	0.624	0.002	0.004	16.298	1.0
	Baseline viral load	0.06	0.004	0.618	-0.299	0.596	18.644	1.0
	Current viral load	0.24	0.060	0.040*	18.366	8.776	-7.033	1.0
	Glucose	0.22	0.048	0.068	1.020	0.550	11.433	1.0
	Triglycerides	0.27	0.070	0.025*	1.974	0.864	13.770	1.0
	Total cholesterol	0.05	0.003	0.671	0.427	1.001	15.129	1.0
	HDL	0.32	0.105	0.006*	-7.309	2.565	27.022	1.0
	LDL	0.06	0.003	0.644	0.564	1.214	15.569	1.0
	IL-6	0.04	0.002	0.733	0.000	0.000	17.638	1.0

*statistically significant at $p < 0.05$

University Malaya

4.4 Non-linear Modelling Analysis Result

ANN models consisting of one input layer, one output layer and one hidden layer topology were constructed with varied different number of hidden neurons in the hidden layer with N_H = 1, 2,.…10. In this study, two separate sets of ANN models were built to explore the role of inflammation marker, IL-6 in determining the body fat composition of MLWH. One set of ANN models were constructed to explore the clinical relevance of BMI and WHR and another two sets of ANN models to explore the accuracy of using the significant variables associated to fat mass in bivariate linear regression analysis. The first four sets used 70% training partition, 20% testing partition and 10% holdout partition while the fifth set used 60% training and 40% testing partition.

1) Set A – ANN models of 16 dependent variables (clinical, metabolic and inflammation variables)
2) Set B – ANN models of 15 dependent variables (clinical and metabolic variables)
3) Set C – ANN models of 2 dependent variables (BMI and WHR)
4) Set D – ANN models of 7 dependent variables (significantly associated with fat mass in bivariate linear regression analysis)
5) Set E – ANN models of 7 dependent variables (significantly associated with fat mass in bivariate linear regression analysis)

The performance of the ANN models was identified through its SSE, RE and MPA% as shown in Table 4.3, Table 4.4, Table 4.5, Table 4.6 and Table 4.7. Figure 4.1 shows SSE against the number of hidden nodes in ANN hidden layer for training models and testing models for Set A. Set A is a set of ANN models built to explore the role of inflammation marker, IL-6 in determining the body fat composition of MLWH by using clinical, metabolic and inflammation variables as its input factor. Based on the Table 4.3, the minimum training SSE obtained is 0.290 and minimum testing SSE obtained is 0.022

Universiti Malaya

while the minimum holdout RE obtained is 0.088. Whereas the average training SSE, average testing SSE and average holdout RE obtained is 0.572, 0.175 and 0.307. The minimum SSE and RE values are smaller when compared with the average SSE and RE values.

Table 4-3 Performance of neural networks using clinical, metabolic and inflammation variables (Set A)

Network model	Number of hidden nodes	SSE_{train}	SSE_{test}	$RE_{holdout}$	MPA%
A1	1	0.581	0.381	0.151	82.47%
A2	2	0.290	0.236	0.387	83.90%
A3	3	0.556	0.322	0.502	80.16%
A4	4	0.362	0.301	0.522	83.00%
A5	5	0.572	0.123	0.315	83.34%
A6	6	0.644	0.027	0.169	82.47%
A7	7	1.020	0.022	0.419	78.27%
A8	8	0.493	0.163	0.088	83.75%
A9	9	0.773	0.071	0.354	82.33%
A10	10	0.432	0.107	0.165	84.84%
Range	Max	1.020	0.381	0.522	84.84%
	Min	0.290	0.022	0.088	78.27%
	Average	0.572	0.175	0.307	82.45%

MPA% value represents the success of the ANN models in solving the problem it was presented, i.e predicting the fat mass, of MLWH using clinical, metabolic and inflammation variables for Set A. The single-hidden layer ANN model with ten hidden neurons in Set A obtained the highest MPA% of 84.84% even though the ANN model with MPA% of 83.90% has the lowest training SSE with two hidden neurons. The model A10 obtained 0.432 SSE for its training subset, 0.107 SSE for its testing error and 0.165 RE error for its holdout subset while model A2 obtained 0.290 SSE for its training subset, 0.236 SSE for its testing error and 0.387 RE error for its holdout subset. The single-hidden layer of ANN model architecture with ten number of hidden neurons in Set A was

Universiti Malaya

accepted as the optimum model because of its highest MPA%.

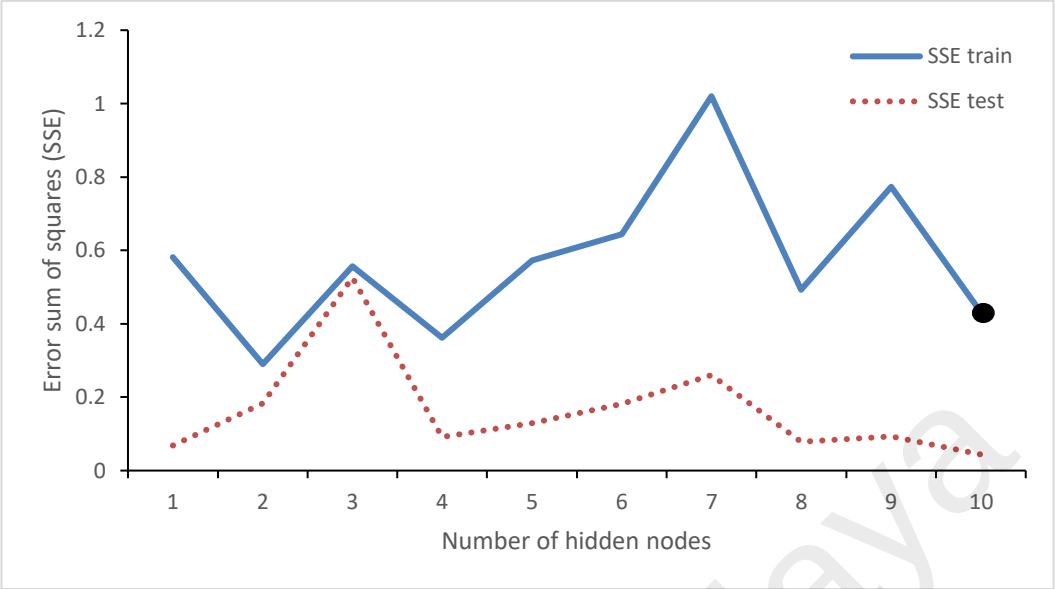

Figure 4-1 Graph of error sum of squares (SSE) against number of hidden nodes in ANN hidden layer for Set A – ANN models with clinical, metabolic and inflammation variables as factor. Straight line represent errors from training subset and dotted line represents error from testing subset. Best model is represented by a dot.

Universiti Malaya

Table 4.4 shows the performance of the ANN models through its SSE, RE and MPA% while Figure 4.2 shows SSE against the number of hidden nodes in ANN hidden layer for training models and testing models for Set B. Set B used clinical and metabolic variables as input factor in modelling the ANN. Based on Table 4.4, the minimum training SSE obtained is 0.443 and minimum testing SSE obtained is 0.044 while the minimum holdout RE obtained is 0.086. Whereas the average training SSE, average testing SSE and average holdout RE obtained is 0.613, 0.170 and 0.307. The minimum SSE and RE values are smaller when compared with the average SSE and RE values in Set B similar with Set A.

Table 4-4 Performance of neural networks using clinical and metabolic variables (Set B)

Network model	Number of hidden nodes	SSE_{train}	SSE_{test}	$RE_{holdout}$	MPA%
B1	1	0.830	0.068	0.179	82.42%
B2	2	0.479	0.183	0.400	83.74%
B3	3	0.511	0.526	0.263	83.36%
B4	4	0.681	0.091	0.086	81.60%
B5	5	0.703	0.129	0.583	82.35%
B6	6	0.577	0.182	0.225	82.10%
B7	7	0.550	0.260	0.112	83.96%
B8	8	0.443	0.079	0.379	83.81%
B9	9	0.720	0.093	0.203	82.23%
B10	10	0.637	0.044	0.643	84.02%
Range	Max	0.830	0.526	0.643	84.02%
	Min	0.443	0.044	0.086	81.60%
	Average	0.613	0.170	0.307	82.96%

Based on Table 4.4, the single-hidden layer ANN model with ten hidden neurons in Set B obtained the highest MPA% of 84.02% even though the ANN model with MPA% of 83.81% has the lowest training SSE with eight hidden neurons. The model B10 obtained 0.637 SSE for its training subset, 0.044 SSE for its testing error and 0.643 RE error for its holdout subset while model B8 obtained 0.443 SSE for its training subset, 0.079 SSE for its testing error and 0.379 RE error for its holdout subset. The single-hidden

Universiti Malaya

layer of ANN model architecture with ten number of hidden neurons in Set B was accepted as the optimum model because of its highest MPA%.

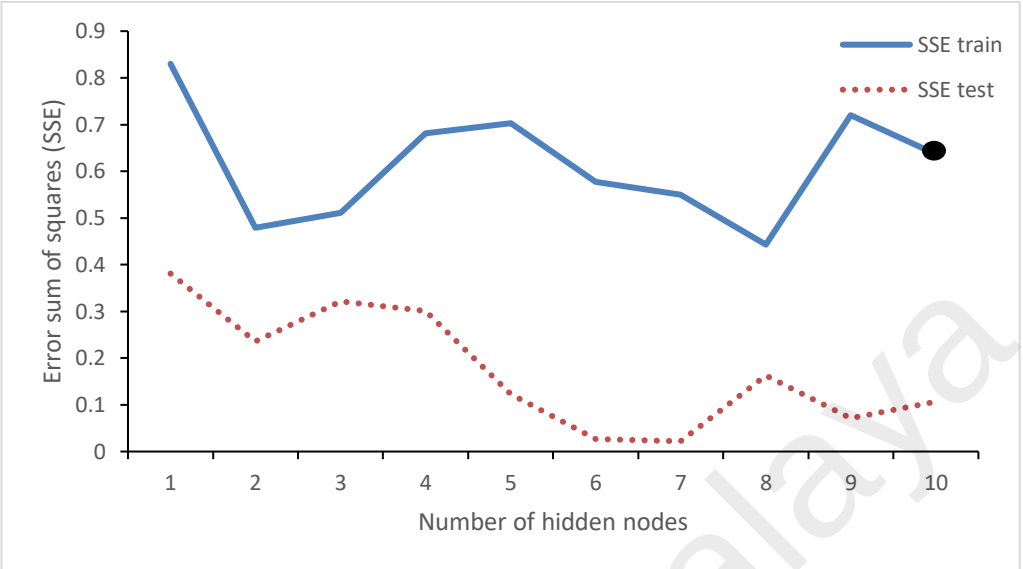

Figure 4-2 Graph of error sum of squares (SSE) against number of hidden nodes in ANN hidden layer for Set B – ANN models with clinical and metabolic variables as factor. Straight line represent errors from training subset and dotted line represents error from testing subset. Best model is represented by a dot.

Universiti Malaya

Table 4.5 shows the performance of the ANN models through its SSE, RE and MPA% while Figure 4.3 shows SSE against the number of hidden nodes in ANN hidden layer for training models and testing models for Set C. Set C used BMI and WHR variables as input factor in modelling the ANN. Based on Table 4.5, the minimum training SSE obtained is 0.993 and minimum testing SSE obtained is 0.019 while the minimum holdout RE obtained is 0.160. Whereas the average training SSE, average testing SSE and average holdout RE obtained is 1.335, 0.298 and 0.473. The minimum SSE and RE values are smaller when compared with the average SSE and RE values in Set C.

Table 4-5 Performance of neural networks using BMI and WHR variables (Set C)

Network model	Number of hidden nodes	SSE_{train}	SSE_{test}	$RE_{holdout}$	MPA%
C1	1	1.183	0.520	0.160	79.22%
C2	2	1.515	0.019	0.366	79.40%
C3	3	1.127	0.421	0.599	78.95%
C4	4	1.479	0.129	0.846	79.72%
C5	5	1.122	0.350	0.850	79.37%
C6	6	1.316	0.383	0.209	79.48%
C7	7	1.344	0.349	0.484	79.25%
C8	8	1.366	0.163	0.251	80.09%
C9	9	1.900	0.155	0.788	73.59%
C10	10	0.993	0.493	0.175	78.81%
Range	Max	1.900	0.520	0.850	80.09%
	Min	0.993	0.019	0.160	73.59%
	Average	1.335	0.298	0.473	78.79%

Based on Table 4.5, the single-hidden layer ANN model with eight hidden neurons in Set C obtained the highest MPA% of 80.09% even though the ANN model with MPA% of 78.81% has the lowest training SSE. The model C8 obtained 1.336 SSE for its training subset, 0.163 SSE for its testing error and 0.251 RE error for its holdout subset while model C10 obtained 0.993 SSE for its training subset, 0.493 SSE for its testing error and 0.175 RE error for its holdout subset. The single-hidden layer of ANN model architecture

Universiti Malaya

with eight number of hidden neurons in Set C was accepted as the optimum model because of its highest MPA%.

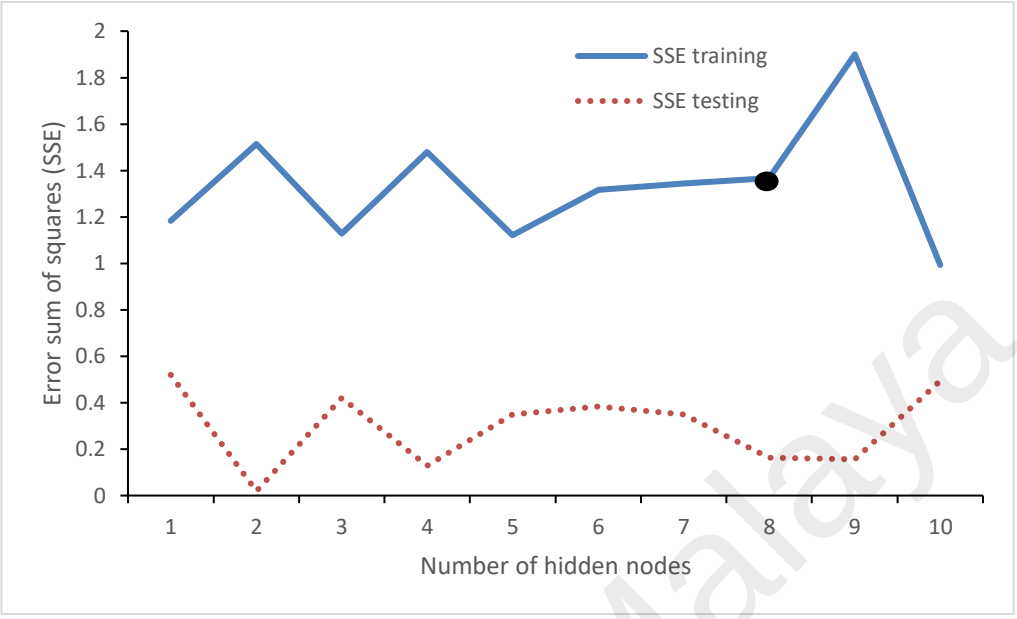

Figure 4-3 Graph of error sum of squares (SSE) against number of hidden nodes in ANN hidden layer for Set C – ANN models with BMI and WHR variables as factor. Straight line represent errors from training subset and dotted line represents error from testing subset. Best model is represented by a dot.

Universiti Malaya

Table 4.6 shows the performance of the ANN models through its SSE, RE and MPA% while Figure 4.4 shows SSE against the number of hidden nodes in ANN hidden layer for training models and testing models for Set D. Set D is a set of ANN models built to explore the accuracy of using the significant variables associated to fat mass in bivariate linear regression analysis. Based on the Table 4.6, the minimum training SSE obtained is 0.722 and minimum testing SSE obtained is 0.006 while the minimum holdout RE obtained is 0.159. Whereas the average training SSE, average testing SSE and average holdout RE obtained is 0.907, 0.111 and 0.569. The minimum SSE and RE values are smaller when compared with the average SSE and RE values.

Table 4-6 Performance of neural networks using significant variables from bivariate linear regression analysis (Set D)

Network model	Number of hidden nodes	SSE_{train}	SSE_{test}	$RE_{holdout}$	MPA%
D1	1	1.115	0.006	0.483	84.29%
D2	2	0.934	0.175	0.235	83.73%
D3	3	1.156	0.036	0.563	83.80%
D4	4	0.756	0.146	0.371	85.26%
D5	5	0.722	0.152	0.380	84.32%
D6	6	0.723	0.218	2.660	85.23%
D7	7	0.801	0.106	0.159	84.85%
D8	8	1.207	0.063	0.440	80.21%
D9	9	0.772	0.131	0.179	84.28%
D10	10	0.887	0.078	0.219	84.62%
Range	Max	1.207	0.218	2.660	85.26%
	Min	0.722	0.006	0.159	80.21%
	Average	0.907	0.111	0.569	84.06%

The single-hidden layer ANN model with four hidden neurons in Set D obtained the highest MPA% of 85.26% even though the ANN model with MPA% of 84.32% has the lowest training SSE. The model D4 obtained 0.756 SSE for its training subset, 0.146 SSE for its testing error and 0.371 RE error for its holdout subset while the model D5 obtained 0.722 SSE for its training subset, 0.152 SSE for its testing error and 0.380 RE error for

Universiti Malaya

its holdout subset. The single-hidden layer of ANN model architecture with four number of hidden neurons in Set D was accepted as the optimum model because of its highest MPA%.

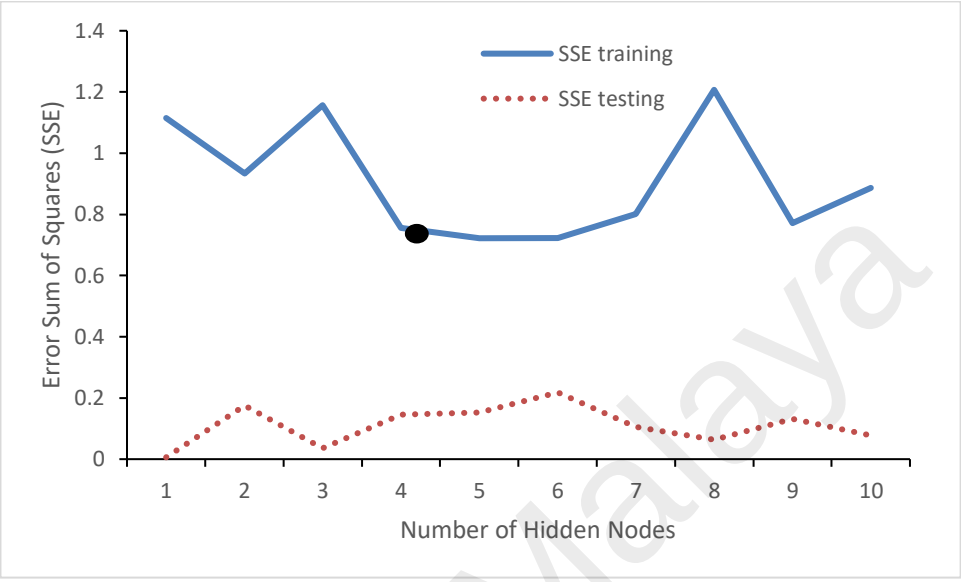

Figure 4-4 Graph of error sum of squares (SSE) against number of hidden nodes in ANN hidden layer for Set D– ANN models using significant variables from bivariate linear regression analysis as factor. Straight line represent errors from training subset and dotted line represents error from testing subset. Best model is represented by a dot.

Universiti Malaya

Table 4.7 shows the performance of the ANN models through its SSE, RE and MPA% while Figure 4.5 shows SSE against the number of hidden nodes in ANN hidden layer for training models and testing models for Set E. Set E is a set of ANN models built to explore the accuracy of using the significant variables associated to fat mass in bivariate linear regression analysis similar with Set D but different in training and testing partition. Based on the Table 4.7, the minimum training SSE obtained is 0.316 and minimum testing SSE obtained is 0.282. Whereas the average training SSE and average testing SSE obtained is 0.593 and 0.792. The minimum SSE and RE values are smaller when compared with the average SSE and RE values.

Table 4-7 Performance of neural networks using significant variables from bivariate linear regression analysis with different training and testing partition (Set E)

Network model	Number of hidden nodes	SSE_{train}	SSE_{test}	MPA%
E1	1	0.362	0.511	85.36%
E2	2	0.496	1.693	85.48%
E3	3	0.316	0.633	85.03%
E4	4	0.388	0.483	85.36%
E5	5	0.447	0.553	83.04%
E6	6	0.932	0.440	83.33%
E7	7	0.569	0.351	84.70%
E8	8	0.757	1.753	83.94%
E9	9	1.066	0.865	83.18%
E10	10	0.530	0.282	84.70%
Range	Max	1.066	1.753	85.48%
	Min	0.316	0.282	83.04%
	Average	0.593	0.792	84.41%

The single-hidden layer ANN model with two hidden neurons in Set E obtained the highest MPA% of 85.48% even though the ANN model with MPA% of 85.03% has the lowest training SSE. The model E2 obtained 0.496 SSE for its training subset and 1.693 SSE for its testing error while the model E3 obtained 0.316 SSE for its training subset and 0.633 SSE for its testing error. The single-hidden layer of ANN model architecture

Universiti Malaya

with two number of hidden neurons in Set E was accepted as the optimum model because of its highest MPA%.

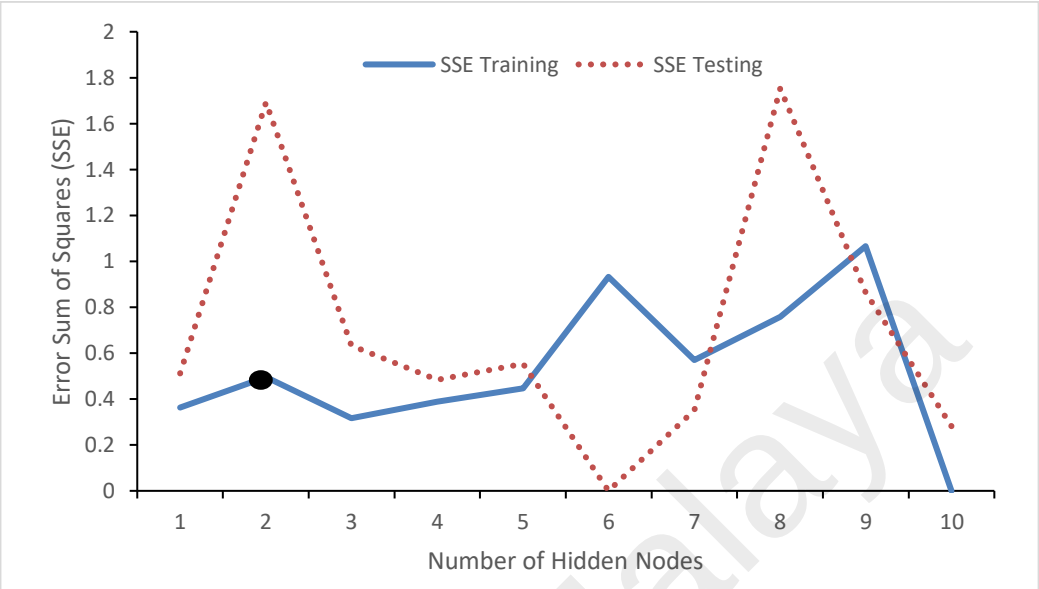

Figure 4-5 Graph of error sum of squares (SSE) against number of hidden nodes in ANN hidden layer for Set E – ANN models using significant variables from bivariate linear regression analysis as factor. Straight line represent errors from training subset and dotted line represents error from testing subset. Best model is represented by a dot.

University Malaya

Table 4.8 shows the comparison of best performed ANN model for each set of ANN models. Set A were developed to predict the FM value by using clinical, metabolic and inflammation components while Set B eliminated the inflammation component and used clinical and metabolic component as input. Set C applies BMI and WHR as its input component to predict the FM value of MLWH. Set D use significant variables associated with fat mass from previous bivariate linear regression analysis. From Table 4.8, it can be noted that the training SSE value for the ANN model (Set A) with the inclusion of inflammation marker as input variable was lesser in comparison to ANN model (Set B) absence of inflammation marker as input variable and the testing SSE value for both ANN models also demonstrates similar result. The holdout RE value for the ANN model (Set A) with the inclusion of inflammation marker as variable was larger albeit small in comparison to ANN model (Set B) absence of IL-6 as input variable. Whereas the training SSE value for the Set C was greater than the training SSE value for the Set A, B, D and E. The testing SSE value for Set C was lesser than the testing SSE value for Set A, B and E. The holdout RE value for Set C was greater than the holdout RE value for the Set A, B and D. The training SSE value for Set D was bigger than the training SSE value for the Set A, B and E although smaller than the training SSE for Set C. The testing SSE value for Set D was the smallest compared to Set A, B, C and E. The holdout RE value for Set D was greater than the holdout RE value for the Set A and B while smaller than Set C. The training SSE value for Set E was smaller than the training SSE value for the Set B, C and D although larger than the training SSE for Set A. The testing SSE value for Set E was the greater compared to Set A, B, C and D. There is no holdout RE for Set E because this set only divide its data into two partitions, training and testing.

Universiti Malaya

Table 4-8 Comparison of best performance ANN model for Set A, B, C, D and E

	Set A	Set B	Set C	Set D	Set E
Minimum SSE					
Training	0.290	0.443	0.993	0.722	0.316
Testing	0.022	0.044	0.019	0.006	0.282
Minimum RE					
Holdout	0.088	0.086	0.160	0.159	-
MPA%	84.84	84.02	80.09	85.26	85.48

MPA% demonstrates that Set E has highest prediction accuracy when compared with Set A, B, C and D. Meanwhile, Set A has higher prediction accuracy when compared with Set B although the difference between Set A and Set B is small which it can be assumed that the accuracy achieved by both ANN sets are similar to each other. This means ANN model performance of Set A and Set B did not differ so much with each other even though models from Set A included inflammation marker as one of its input components. However, Set C has low prediction compared to Set A, B, D and E.

In order to identify the significance of IL-6 in predicting the body fat composition of HIV-positive males , independent sample *t*-test was conducted on Set A and Set B and it is found that no significant difference reported in MPA% between Set A and Set B where $t(13) = 0.75$, $p<.05$, 95% *C.I.* [-0.95% – 1.96%] even though the Set B had slightly higher MPA% ($M = 82.96\%$, $SD = 1.92\%$) as compared to Set A ($M = 82.45\%$, $SD = 0.91\%$) as shown in Table 4.9.

Table 4-9 Independent sample *t*-test conducted to test the prediction accuracy of fat mass in MLHIV using average of MPA%

	Set A– with IL-6 (n-10)	Set B– without IL-6 (n=10)
	$t(13) = 0.75$, $p < .05$	
Mean (%)	82.45	82.96
SD (%)	0.91	1.92
Implication	No significant difference reported in MPA% between Set A and Set B ANN models. IL-6 role in predicting fat mass in MLHIV was not significance.	

Set A: ANN models using clinical, metabolic and inflammation variables, Set B: ANN models using clinical and metabolic variables.

Universiti Malaya

A one-way between sets ANOVA was conducted to compare the prediction accuracy on fat mass in Set A, B, C, D and E. There was a statistically significant difference in prediction accuracy on fat mass in MLWH at the p <.05 level for the five ANN models sets [F (4, 45) = 8.802, p = 0.00] as shown in Table 4.10.

Table 4-10 One-way ANOVA result among Set A, B, C, D and E

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	0.455	4	0.114	8.802	0.000*
Within Groups	0.582	45	0.013		
Total	1.037	49			

*statistically significant at p < 0.05

Post hoc comparisons using the Tukey HSD test in Table 4.11 indicated that the prediction accuracy mean score for the Set C (M = 78.79%, SD = 1.86%, p = 0.000) was statistically significantly lower than Set A, B and D. However, the prediction accuracy mean score for Set A (M = 82.45%, SD = 1.92%) was not statistically significant difference with Set B (M = 82.96%, SD = 0.91%, p = 0.940) and Set D (M = 84.06%, SD = 1.45%, p = 0.129). There was also no statistically significant difference mean score prediction accuracy between Set B and Set D (p = 0.470) and Set E (p = 0.203). The prediction accuracy mean score for Set E (M = 84.41%, SD = 0.96%) was statistically significant difference with Set A (p = 0.038) and Set D (p = 0.00). These results suggest that there the prediction accuracy of every set of ANN models is similar with each other except for Set C and Set E. It is also found that the result for Set A and Set B ANN models is similar with independent sample *t*-test conducted previously where the prediction accuracy on fat mass for these two sets is considered statistically similar with each other.

Universiti Malaya

Table 4-11 Post hoc comparisons using Tukey HSD test among Set A, B, C, D and E.

Set	Set	Mean Difference (%)	Sig.
A	B	-0.005	0.940
	C	0.037*	0.000
	D	-0.016	0.129
	E	-0.020*	0.038
B	A	0.005	0.940
	C	0.042*	0.000
	D	-0.011	0.470
	E	-0.015	0.203
C	A	-0.037*	0.000
	B	-0.042*	0.000
	D	-0.053*	0.000
	E	-0.056*	0.000
D	A	0.016	0.129
	B	0.011	0.470
	C	0.053*	0.000
	E	-0.004	0.984
E	A	0.020*	0.038
	B	0.015	0.203
	C	0.056*	0.000
	D	0.004	0.984

* The mean difference is significant at the 0.05 level.

4.5 Summary

The analysis of relationship between FM and inflammation, metabolic and clinical risk factors using ANN was presented in this chapter. To find the significant impact the inflammation variable has on FM in MLWH, two set of ANN models were trained and compared. One set (Set A) used inflammation, metabolic and clinical variables as input while the other ANN set (Set B) remove inflammation variable from its network input and trained only metabolic and clinical variables. No significant difference in prediction accuracy was found between Set A and Set B from independent sample *t*-test analysis.

Universiti Malaya

Apart from that, one set of ANN models was trained by using BMI and WHR as its network input and it was found that its performance had the lowest prediction accuracy among other ANN sets.

From the result acquired from linear regression analysis, the variables that were significantly associated with FM were used as network input in Set D and Set E. Both obtained higher prediction accuracy compared to Set A, B and C, however, both sets had no significant difference in their prediction performance as reported in post hoc comparison test.

Universiti Malaya

CHAPTER 5: DISCUSSION

Overall, the results demonstrated that ANN can be used as a tool to predict the relationship between the FM of MLWH with the clinical, metabolic and inflammation components. The development of the ANN is considered important because of its ability to improve the efficiency and efficacy of predicting the body fat composition and manage suitable interventions in MLWH. By predicting the FM, the body fat composition of MLWH can be altered by clinically monitoring the risk factors especially through the significant risk factors associated with the FM as determined by bivariate linear regression test in this study (weight, WHR, BMI, triglycerides, HDL and HIV viral load value) to best estimate the degree of change needed to reach the target healthy FM for MLWH. The significant association of these factors suggest the best clinical measure to control in comparison to the other risk factors. The ANN models that apply these parameters in its network, Set D and Set E also shown increased prediction accuracy on FM in comparison to the other ANN sets. ANN model from Set E tops other ANN sets where its best performing ANN model achieve 85.26% MPA% with four hidden nodes in its hidden layer. Furthermore, employment of ANN prediction reflects faster change and faster accurate assessment of FM value when the risk factors are monitored and controlled in MLWH using timely interventions such as in medication therapy, additional testing, and medical treatment.

One area required focus in particular was the presence of inflammation marker, IL-6 as one of risk factors in determining overall body fat composition in MLWH. Nevertheless, there is no change of the percentage of MPA% for FM prediction if inflammation marker, IL-6 was included as input variable with 82.45% average MPA% in Set A compared to ANN models without IL-6 included as input variable with 82.96% average MPA% as in Set B. However, for these two sets of ANN models, the best

Universiti Malaya

performing ANN model with 84.84% of accuracy is the ANN model that uses independents clinical, metabolic and inflammation variables and ten hidden nodes in its single-hidden layer. Inflammation factor which in this case, IL-6 is associated with body fat composition in PLWH and impacts the metabolic indicators and body fat percentage as well as inflammation responses and it is predicted to influence the model by raising the percentage of prediction accuracy. IL-6 is associated to the body fat composition of MLWH by having inter-relationship and complex interaction starting from the generation of inflammation mediator cytokines which the IL-6 is one of the participants where it is either secreted by adipose tissues or present in body liquid plasma. It then becomes the changes influencing the fat morphology, adipocyte differentiation and lipid profile infecting the PLWH (Cervia et al., 2010). In this aspect, the IL-6 is associated to the body fat composition in PLWH and impact the metabolic indicators and body fat percentage as well as inflammation responses. However, the performance of ANN depends on the predictive efficiency of input variables and the correlation level that exist for each input variables with the output variable (Bishop, 1995) even though inflammation marker is one of the factors influencing the body fat composition in MLWH as the interaction of the immune system and metabolic pathway trigger components associated with fat redistribution and metabolic disturbance (Koethe et al., 2013). Based on the result, the inclusion of IL-6 as inflammation factor may not be sufficient and warrants additional inflammation marker such as CRP, TNF-α, D-dimer and soluble CD14. Other components such as microbial translocation from the gut, changes in gut microbiome (Tenorio et al., 2014) and co-infection with latent viruses like human herpes viruses (Munawwar & Singh, 2016) should be considered and included where these factors have been shown to influence the IL-6 value in MLWH.

The use of clinical and metabolic factors to predict the FM in MLWH shown high degree of accuracy accounting for 84.02% MPA% as achieved by ANN model with ten

Universiti Malaya

hidden nodes in hidden layer (Model B10). It is known that metabolic parameters – glucose, triglycerides, HDL, LDL and total cholesterol is associated in body fat composition of MLWH. Even though the pathogenesis of body fat and metabolic changes occurring in PLWH is not completely clear, the explanation is that the metabolic irregularity is caused by changes in body fat distribution especially in visceral fat accumulation and antiretroviral treatment (Hejazi & Rajikan, 2015). It is on this account the metabolic factor is highly considered as significant in modulating FM and in overall, the body composition development in PLWH. Consumption of lipid-lowering agent, for example statins usually is taken alongside ART regimens as management measures for metabolic syndrome in PLWH in addition to healthy lifestyle choice (Calza et al., 2016). The clinical factor – age, height, weight, WHR, BMI, administered duration of cART by months, baseline CD4 cell count, CD4 cell count, baseline $\log_{10}$ HIV viral load and $\log_{10}$ HIV viral load in this case provided separate but mutual contribution in development of ANN models. The importance of clinical factor is critical in the sense of providing clinical assessment through external measurement or biochemical analysis of blood sample to assess prevalence of diseases. It is of considerable importance to find how well the ANN model responds to only inclusion of metabolic influence in enhancing the ANN learning process.

Set C is developed to investigate the anthropometric indicators to effectively predict the body composition that is FM in MLWH and BMI and WHR have been shown to be associated with body fat composition. From the clinical perspective, both parameters are good indicator to assess body fat composition even though it is unable to distinguish apart the abdominal subcutaneous fat, total abdominal fat and total body fat (Beraldo et al., 2016) in addition to BMI and WHR high correlation to each other from the statistical view. Set C has low mean prediction compared to other ANN sets where it did not reach 80% mean prediction accuracy. This is because the parameter variables for Set C is only

Universiti Malaya

BMI and WHR and it is not sufficient in training ANN model although BMI and WHR are good indicator to assess body fat composition where more components are required for the model to learn effectively in order to generate correct desired output. Additional features from demographic and clinical risk factors add to learning accuracy as observed from Set A and Set B where it improves the training and learning process for ANN model.

Different combination of split ratio dataset has been investigated with only the risk factors that are significantly associated with FM. The reason is that the Set D has obtained highest accuracy compared to Set A, B and C and hence, Set E was developed to assess the network prediction accuracy but with different split ratio. The comparison of the predictability of Set D (70% training, 20% testing, 10% holdout) and Set E (60% training, 40% testing) shows higher predictive capacity of 60% training and 40% testing split ratio dataset. The analysis also confirms that by using the variables significantly associated with FM, albeit different split ratio, higher accuracy is achieved compared to other network models that use inflammation, metabolic and clinical variables in its network. However, the ANN performance of both sets that use risk factors significantly associated with FM as input variable remains similar with each other. No statistically significant difference is found between Set D and Set E. It can be implied that there is no clear relationship between the proportion data used for training, testing and holdout dataset with ANN model performance.

In context of ANN topology, number of hidden neurons in hidden layer displays important criteria in finding ANN optimal solution performance and it necessitates the network architecture to be carefully selected to secure the correct respond behavior. In this study, systematic experimentation (Sheela & Deepa, 2013) is applied to discover the best working system for the dataset because it is not possible analytically to calculate the number of nodes for hidden layer, hence the need to specify number of hidden nodes in

Universiti Malaya

hidden layer. Ten ANN models for each Set A, B, C, D and E were built to avoid the model from encountering local minima in their search of least function error and to find the point that minimizes the value of the function. Incremental improvement is executed on the model to find the global minimum instead of local minimum where each step of improvement leads to the least error for entire function domain. Findings reported in the literature is a good starting point to find the number of hidden nodes on instances of prediction problem similar to the prediction problem at hand (Zhang, 2006). By finding the configuration of neural networks used and by testing the transferability of the model hyperparameters, optimal ANN model algorithm can be correctly identified for the research application in this study.

The training sample used in this study were relatively small. Due to the nature of study, the data gathering of MLWH was incomplete and the collection of informative data was insufficient making the sample size of MLWH small. Ideally, training sample should be large enough to provide sufficient training data for the model learning process to cover training subset, test subset and holdout subset. Large training data means increase flexibility and power for the ANN algorithm to learn complex non-linear relationship between input and output items, however the same can be achieved for small dataset analysis (Pasini, 2015). By increasing the size and diversity of the training database, it is believed the ANN model performance can improve its accuracy and generalizability. One of the advantages of ANN modelling is it performs its learning process in the background by its own without external supervision and it greatly reduce the need for individualized model regulation. However, owing to the same feature, the structure of the models produced are not amenable to inspection and it is up to the network system entirely to self-evaluate its network behavior and arrive at conclusion with its prediction algorithm uninterpretable due to its deep learning characteristic (Tan et al., 2015).

Universiti Malaya

One potential limitation of this study is that it was not developed and tested on a generalized PLWH population, but rather defined population of MLWH. By addressing MLWH cohort only, this study has overlooked the female cohort and this distinction is highlighted to stress the work of study may represent only specific segment of PLWH population. This is because of the small number of WLWH cohort in this study which will introduce unintended gender bias due to skewed distribution of sex ratio in the entire cohort. Another formal study is required to develop overall ANN prediction models for the entire PLWH population and determine the model performance in predicting the body FM. However, development of separate model for separate gender or specific PLWH population should also be considered as the performance of the prediction models may be improved depending on certain PLWH population.

Another aspect of limitation in this study is no variables concerning the type of ART drugs are employed in the prediction model despite ART being a strong factor influencing the development of DM, dyslipidemia and lipodystrophy syndrome as discussed in the literature review. The variable is not included due to its complex relationship with other predictor variables and in addition to the variable categorized as categorical data. The ANN used in this study is not diverse enough to handle both categorical and continuous variables in its network. Further complex deep learning and extensive training is required for the network to produce accurate prediction model. Maybe other learning method, for example random decision forest can be employed instead of ANN where it can incorporate type of ART drugs used as predictor in future analysis.

Universiti Malaya

CHAPTER 6: CONCLUSION AND FUTURE WORK

6.1 Summary

In this study, the role of IL-6 which is a pro-inflammatory marker has been investigated among the clinical, metabolic and FM components in MLWH. The relationship established using ANN as prediction analysis tool signifies the contribution of IL-6 (or the lack of) in predicting the FM value of MLWH. From the study finding, it is discovered that IL-6 did not influence the prediction analysis for FM. Both ANN models with and without IL-6 show over 80% accuracy, thus IL-6 does not have significant role in the prediction. Compared to its counterpart, the ANN model with clinical and metabolic components without the inflammation component, the function error obtained from training between these two models were almost identical. When *t*-test were conducted between these ANN models, no significant difference was found, and this further confirmed our initial observation that IL-6 had no clear influence on ANN prediction analysis of FM in MLWH.

In addition, the high prediction accuracy (over 80%) of FM in prediction models that use inflammation, metabolic and clinical factors as input indicates the success of ANN in predicting the fat mass in MLWH. Similar result also was reached by prediction models that used only metabolic and clinical factor as input, further shows that metabolic and clinical factors are also important in the prediction models. This prediction success of ANN may owe to its flexible and adaptive non-linear modelling factor because ANN incorporates deep learning algorithm to model complex relationship pattern that exists between data.

Based on the linear regression analysis, BMI and WHR alone are significant predictors for FM even though the ANN models using BMI and WHR has low accuracy when compared with other set of ANN models. The highest accuracy it can achieved as

Universiti Malaya

demonstrated by the prediction model was only at 78.81%. However, the prediction accuracy had shown significant increase when both variables, BMI and WHR were used with other variables. Moreover, the prediction analysis done on significant parameters associated with FM only as model input have shown high prediction accuracy and surpassed prediction performance of other models. Based on this finding, it is stipulated that an alteration of the predictor parameters is able to adjust the FM in MLWH.

The anticipated result of ANN prediction analysis on FM measurement in MLWH play an important role in clinicians decision-making related to adopting novel and innovative machine learning based techniques in HIV field. The clinical significance of this study is that early diagnosis and prognosis can be expected without the need of DEXA technology to assist clinicians in assessing health information of their patient. The high accuracy of ANN methodology could rival the DEXA examination in future as ANN prediction of FM able to predict more than 80% accuracy although the current score could be improved with more data and training.

ANN offers accurate and efficient FM assessment and prediction of clinical status in MLWH with related-metabolic syndromes and through the method, efficient clinical monitoring of the risk factors can be supervised mainly the weight, WHR, BMI, triglycerides, HDL and HIV viral load value. In a way, FM in MLWH can be controlled through these risk factors management by employing clinical prevention measure of consuming lipid-lowering agent, for example statins alongside cART treatment together with healthy lifestyle choice. The accuracy of these prediction models could be further tested in interventional studies.

Universiti Malaya

6.2 Limitation

The training sample used in this study were relatively small. Due to the nature of study, the data gathering of MLWH was incomplete and the collection of informative data was insufficient making the sample size of MLWH small. Ideally, training sample should be large enough to provide sufficient training data for the model learning process to cover training subset, test subset and holdout subset. By increasing the size and diversity of the training database, it is believed the ANN model performance can improve its accuracy and generalizability. One of the advantages of ANN modelling is it performs its learning process in the background by its own without external supervision and it greatly reduce the need for individualized model regulation. However, owing to the same feature, the structure of the models produced are not amenable to inspection and it is up to the network system entirely to self-evaluate its network behavior and arrive at conclusion with its prediction algorithm uninterpretable due to its deep learning characteristic.

6.3 Future Work

Future work conducted will be focused on deeper analysis of ANN algorithm, new proposals to try different methods or work based on curiosity. This thesis has been mainly focused on MLWH cohort as discussed in Chapter 5 where no WLWH cohort is represented. Future analysis could be conducted based on entire HIV population independent of gender using different types of machine learning for example support vector machine, random forest or boosted decision tree. It is interesting to see how these different approaches work in comparison to each other. Each algorithm would for instance aid to distinguish the complex problem in their own separate way. In addition, additional risk factor for example, the influence of lipid lowering-agent in assessing and predicting the FM of these patients could be investigated since they help in controlling lipid levels. Besides, the other inflammatory biomarkers such as CRP, TNF-α, D-dimer and soluble

Universiti Malaya

CD14 could also be incorporated into future analysis. With these multi data, the efficiency of ANN model performance could potentially be improved, and its accuracy increased.

Universiti Malaya

REFERENCES

Aberg, J. A. (2009). Lipid management in patients who have HIV and are receiving HIV therapy. *Endocrinology and Metabolism Clinics of North America*, *38*(1), 207–222.

Aberg, J. A. (2012). Aging, inflammation, and HIV infection. *Topics in Antiviral Medicine*, *20*(3), 101.

Adams, S. T., & Leveson, S. H. (2012). Clinical prediction rules. *Bmj*, *344*, d8312.

Adewole, O. O., Eze, S., Betiku, Y. E., Anteyi, E., Wada, I., Ajuwon, Z., & Erhabor, G. (2010). Lipid profile in HIV/AIDS patients in Nigeria. *African Health Sciences*, *10*(2).

Alencastro, P. R., Barcellos, N. T., Wolff, F. H., Ikeda, M. L. R., Schuelter-Trevisol, F., Brandão, A. B. M., & Fuchs, S. C. (2017). People living with HIV on ART have accurate perception of lipodystrophy signs: a cross-sectional study. *BMC Research Notes*, *10*(1), 1–8.

Andany, N., Raboud, J. M., Walmsley, S., Diong, C., Rourke, S. B., Rueda, S., Rachlis, A., Wobeser, W., MacArthur, R. D., & Binder, L. (2011). Ethnicity and gender differences in lipodystrophy of HIV-positive individuals taking antiretroviral therapy in Ontario, Canada. *HIV Clinical Trials*, *12*(2), 89–103.

Anglemyer, A., Rutherford, G. W., Easterbrook, P. J., Horvath, T., Vitoria, M., Jan, M., & Doherty, M. C. (2014). Early initiation of antiretroviral therapy in HIV-infected adults and adolescents: a systematic review. *Aids*, *28*, S105–S118.

Aounallah, M., Dagenais-Lussier, X., El-Far, M., Mehraj, V., Jenabian, M.-A., Routy, J.-P., & van Grevenynghe, J. (2016). Current topics in HIV pathogenesis, part 2: Inflammation drives a Warburg-like effect on the metabolism of HIV-infected subjects.

Universiti Malaya

Cytokine & Growth Factor Reviews, *28*, 1–10.

Appay, V., & Sauce, D. (2008). Immune activation and inflammation in HIV-1 infection: causes and consequences. *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland*, *214*(2), 231–241.

Armstrong, C., Liu, E., Grinspoon, S., Okuma, J., Spiegelman, D., Guerino, C., Njelekela, M., Fawzi, W., & Hawkins, C. (2011). Dyslipidemia in an HIV-positive, antiretroviral treatment-naïve population in Dar es Salaam, Tanzania. *Journal of Acquired Immune Deficiency Syndromes (1999)*, *57*(2), 141.

Barnes, T. C., Anderson, M. E., & Moots, R. J. (2011). The many faces of interleukin-6: the role of IL-6 in inflammation, vasculopathy, and fibrosis in systemic sclerosis. *International Journal of Rheumatology*, *2011*.

Baughman, D. R., & Liu, Y. A. (2014). *Neural networks in bioprocessing and chemical engineering*. Academic press.

Bazzocchi, A., Ponti, F., Albisinni, U., Battista, G., & Guglielmi, G. (2016). DXA: technical aspects and application. *European Journal of Radiology*, *85*(8), 1481–1492.

Beraldo, R. A., Meliscki, G. C., Silva, B. R., Navarro, A. M., Bollela, V. R., Schmidt, A., & Foss-Freitas, M. C. (2016). Comparing the ability of anthropometric indicators in identifying metabolic syndrome in HIV patients. *PLoS One*, *11*(2), e0149905.

Beraldo, R. A., Vassimon, H. S., Navarro, A. M., & Foss-Freitas, M. C. (2015). Development of predictive equations for total and segmental body fat in HIV-seropositive patients. *Nutrition*, *31*(1), 127–131. https://doi.org/10.1016/j.nut.2014.05.013

Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford University Press.

Universiti Malaya

Bonet, I., García, M. M., Saeys, Y., Van de Peer, Y., & Grau, R. (2007). Predicting Human Immunodeficiency Virus (HIV) drug resistance using recurrent neural networks. *International Work-Conference on the Interplay Between Natural and Artificial Computation*, 234–243.

Botros, D., Somarriba, G., Neri, D., & Miller, T. L. (2012). Interventions to address chronic disease and HIV: strategies to promote exercise and nutrition among HIV-infected individuals. *Current Hiv/Aids Reports*, *9*(4), 351–363.

Bourgi, K., Wanjalla, C., & Koethe, J. R. (2018). Inflammation and metabolic complications in HIV. *Current HIV/AIDS Reports*, *15*(5), 371–381.

Brown, T. T., Chen, Y., Currier, J. S., Ribaudo, H. J., Rothenberg, J., Dub??, M. P., Murphy, R., Stein, J. H., & McComsey, G. A. (2013). Body composition, soluble markers of inflammation, and bone mineral density in antiretroviral therapy-Naive HIV-1-infected individuals. *Journal of Acquired Immune Deficiency Syndromes*, *63*(3), 323–330. https://doi.org/10.1097/QAI.0b013e318295eb1d

Brown, T. T., Cole, S. R., Li, X., Kingsley, L. A., Palella, F. J., Riddler, S. A., Visscher, B. R., Margolick, J. B., & Dobs, A. S. (2005). Antiretroviral therapy and the prevalence and incidence of diabetes mellitus in the multicenter AIDS cohort study. *Archives of Internal Medicine*, *165*(10), 1179–1184. https://doi.org/10.1001/archinte.165.10.1179

Brown, T. T., Tassiopoulos, K., Bosch, R. J., Shikuma, C., & McComsey, G. A. (2010). Association between systemic inflammation and incident diabetes in HIV-infected patients after initiation of antiretroviral therapy. *Diabetes Care*, *33*(10), 2244–2249.

Brownbill, R. A., & Ilich, J. Z. (2005). Measuring body composition in overweight individuals by dual energy x-ray absorptiometry. *BMC Medical Imaging*, *5*(1), 1.

Universiti Malaya

Bruchfeld, J., Correia-Neves, M., & Källenius, G. (2015). Tuberculosis and HIV coinfection. *Cold Spring Harbor Perspectives in Medicine*, *5*(7), a017871.

Burton, R. F. (2017). Relationships among fat mass, fat-free mass and height in adults: A new method of statistical analysis applied to NHANES data. *American Journal of Human Biology*, *29*(3), e22941.

Calza, L., Colangeli, V., Manfredi, R., Bon, I., Re, M. C., & Viale, P. (2016). Clinical management of dyslipidaemia associated with combination antiretroviral therapy in HIV-infected patients. *Journal of Antimicrobial Chemotherapy*, *71*(6), 1451–1465.

Caron-Debarle, M., Lagathu, C., Boccara, F., Vigouroux, C., & Capeau, J. (2010). HIV-associated lipodystrophy: from fat injury to premature aging. *Trends in Molecular Medicine*, *16*(5), 218–229. https://doi.org/10.1016/j.molmed.2010.03.002

Cavalcanti, R. B., Cheung, A. M., Raboud, J., & Walmsley, S. (2005). Reproducibility of DXA estimations of body fat in HIV lipodystrophy: implications for clinical research. *Journal of Clinical Densitometry*, *8*(3), 293–297.

Cervia, J. S., Chantry, C. J., Hughes, M. D., Alvero, C., Meyer 3rd, W. A., Hodge, J., Borum, P., Moye Jr, J., Spector, S. A., & Team, P. 1010. (2010). Associations of proinflammatory cytokine levels with lipid profiles, growth, and body composition in HIV-infected children initiating or changing antiretroviral therapy. *The Pediatric Infectious Disease Journal*, *29*(12), 1118–1122. https://doi.org/10.1097/INF.0b013e3181ed9f4c

Chen, J., Hartono, J. R., John, R., Bennett, M., Zhou, X. J., Wang, Y., Wu, Q., Winterberg, P. D., Nagami, G. T., & Lu, C. Y. (2011). Early interleukin 6 production by leukocytes during ischemic acute kidney injury is regulated by TLR4. *Kidney*

Universiti Malaya

International, *80*(5), 504–515.

Chor, L. (2015). *Fitness Fad Fridays: Analysing body fat with scary precision using a DEXA scan*. Coconuts Hong Kong. https://coconuts.co/hongkong/news/fitness-fad-fridays-analysing-body-fat-scary-precision-using-dexa-scan/

Curran, A., & Ribera, E. (2011). From old to new nucleoside reverse transcriptase inhibitors: changes in body fat composition, metabolic parameters and mitochondrial toxicity after the switch from thymidine analogs to tenofovir or abacavir. *Expert Opinion on Drug Safety*, *10*(3), 389–406.

De Clercq, E., & Herdewijn, P. (2010). Strategies in the design of antiviral drugs. *Pharmaceutical Sciences Encyclopedia: Drug Discovery, Development, and Manufacturing*, 1–56.

Deeb, O., & Jawabreh, M. (2011). *Exploring QSARs for Inhibitory Activity of Cyclic Urea and Nonpeptide-Cyclic Cyanoguanidine Derivatives HIV-1 Protease Inhibitors by Artificial Neural Network*.

Delpierre, C., Bonnet, E., Marion-Latard, F., Aquilina, C., Obadia, M., Marchou, B., Massip, P., Perret, B., & Bernard, J. (2007). Impact of HIV infection on total body composition in treatment—Naive men evaluated by dual-energy X-ray absorptiometry comparison of 90 untreated HIV-infected men to 241 controls. *Journal of Clinical Densitometry*, *10*(4), 376–380.

Desai, D. V., & Kulkarni, S. S. (2015). Herpes simplex virus: the interplay between HSV, host, and HIV-1. *Viral Immunology*, *28*(10), 546–555.

Dkhichi, F., & Oukarfi, B. (2014). Neural Network Training By Gradient Descent Algorithms: Application on the Solar Cell. *International Journal of Innovative Research*

Universiti Malaya

in Science Engineering and Technology, *3*, 15696–15702. https://doi.org/10.15680/IJIRSET.2014.0308084

Doitsh, G., & Greene, W. C. (2016). Dissecting how CD4 T cells are lost during HIV infection. *Cell Host & Microbe*, *19*(3), 280–291.

Dorr, P., Westby, M., Dobbs, S., Griffin, P., Irvine, B., Macartney, M., Mori, J., Rickett, G., Smith-Burchnell, C., & Napier, C. (2005). Maraviroc (UK-427,857), a potent, orally bioavailable, and selective small-molecule inhibitor of chemokine receptor CCR5 with broad-spectrum anti-human immunodeficiency virus type 1 activity. *Antimicrobial Agents and Chemotherapy*, *49*(11), 4721–4732.

Drăghici, S., & Potter, R. B. (2003). Predicting HIV drug resistance with neural networks. *Bioinformatics*, *19*(1), 98–107. https://doi.org/10.1093/bioinformatics/19.1.98

Dreiseitl, S., & Ohno-Machado, L. (2002). Logistic regression and artificial neural network classification models: a methodology review. *Journal of Biomedical Informatics*, *35*(5–6), 352–359.

Earthman, C. P., Matthie, J. R., Reid, P. M., Harper, I. T., Ravussin, E., & Howell, W. H. (2000). A comparison of bioimpedance methods for detection of body cell mass change in HIV infection. *Journal of Applied Physiology*, *88*(3), 944–956. https://doi.org/10.1152/jappl.2000.88.3.944

Echavez, M. (2005). Relationship between lipoatrophy and quality of life. *The AIDS Reader*, 369.

El-Mikkawy, D. M. E., EL-Sadek, M. A., EL-Badawy, M. A., & Samaha, D. (2020). Circulating level of interleukin-6 in relation to body mass indices and lipid profile in Egyptian adults with overweight and obesity. *Egyptian Rheumatology and Rehabilitation*,

Universiti Malaya

47(1), 1–7.

Eswaran, C., & Logeswaran, R. (2011). Improved Adaptive Neuro-Fuzzy Inference System for HIV/AIDS Time Series Prediction. *International Conference on Informatics Engineering and Information Science*, 1–13.

Falutz, J. (2011). Management of fat accumulation in patients with HIV infection. *Current HIV/AIDS Reports*, *8*(3), 200–208.

Fauci, A. S., & Folkers, G. K. (2012). Toward an AIDS-free generation. *Jama*, *308*(4), 343–344.

Finkelstein, J. L., Gala, P., Rochford, R., Glesby, M. J., & Mehta, S. (2015). HIV/AIDS and lipodystrophy: implications for clinical management in resource-limited settings. *Journal of the International AIDS Society*, *18*(1), 19033.

Fogel, G. B., Lamers, S. L., Liu, E. S., Salemi, M., & McGrath, M. S. (2015). Identification of dual-tropic HIV-1 using evolved neural networks. *Biosystems*, *137*, 12–19.

Forrester, J. E., Spiegelman, D., Tchetgen, E., Knox, T. A., & Gorbach, S. L. (2002). Weight loss and body-composition changes in men and women infected with HIV. *The American Journal of Clinical Nutrition*, *76*(6), 1428–1434.

Fuentes, E., Fuentes, F., Vilahur, G., Badimon, L., & Palomo, I. (2013). Mechanisms of chronic state of inflammation as mediators that link obese adipose tissue and metabolic syndrome. *Mediators of Inflammation*, *2013*.

Funderburg, N. T., & Mehta, N. N. (2016). Lipid Abnormalities and Inflammation in HIV Inflection. *Current HIV/AIDS Reports*, *13*(4), 218–225. https://doi.org/10.1007/s11904-

Universiti Malaya

016-0321-0

Gabay, C. (2006). Interleukin-6 and chronic inflammation. *Arthritis Research & Therapy*, *8*(2), S3.

Gaj, S., Yang, M., Nakamura, K., & Li, X. (2020). Automated cartilage and meniscus segmentation of knee MRI with conditional generative adversarial networks. *Magnetic Resonance in Medicine*, *84*(1), 437–449.

Gao, S., Durstine, J. L., Koh, H.-J., Carver, W. E., Frizzell, N., & Carson, J. A. (2017). Acute myotube protein synthesis regulation by IL-6-related cytokines. *American Journal of Physiology-Cell Physiology*, *313*(5), C487–C500.

Goodsitt, M. M. (1992). Evaluation of a new set of calibration standards for the measurement of fat content via DPA and DXA. *Medical Physics*, *19*(1), 35–44.

Gottlieb, M. S., Schroff, R., Schanker, H. M., Weisman, J. D., Fan, P. T., Wolf, R. A., & Saxon, A. (1981). *Pneumocystis carinii* Pneumonia and Mucosal Candidiasis in Previously Healthy Homosexual Men. *New England Journal of Medicine*, *305*(24), 1425–1431. https://doi.org/10.1056/NEJM198112103052401

Grant, P. M., Kitch, D., McComsey, G. A., Collier, A. C., Bartali, B., Koletar, S. L., Erlandson, K. M., Lake, J. E., Yin, M. T., & Melbourne, K. (2016). Long-term body composition changes in antiretroviral-treated HIV-infected individuals. *AIDS (London, England)*, *30*(18), 2805.

Greenberg, M., Cammack, N., Salgo, M., & Smiley, L. (2004). HIV fusion and its inhibition in antiretroviral therapy. *Reviews in Medical Virology*, *14*(5), 321–337.

Grunfeld, C. (2010). Dyslipidemia and its treatment in HIV infection. *Topics in HIV*

Universiti Malaya

Medicine: A Publication of the International AIDS Society, USA, *18*(3), 112.

Gupta, T. (2017). *Deep Learning: Feedforward Neural Network*. Medium: Towards Data Science. https://towardsdatascience.com/deep-learning-feedforward-neural-network-26a6705dbdc7

Gutierrez, D. A., Puglisi, M. J., & Hasty, A. H. (2009). Impact of increased adipose tissue mass on inflammation, insulin resistance, and dyslipidemia. *Current Diabetes Reports*, *9*(1), 26–32.

Hadigan, C., & Kattakuzhy, S. (2014). Diabetes mellitus type 2 and abnormal glucose metabolism in the setting of human immunodeficiency virus. *Endocrinology and Metabolism Clinics*, *43*(3), 685–696.

Hailu, T. G. (2015). Comparing data mining techniques in HIV testing prediction. *Intelligent Information Management*, *7*(03), 153.

Hammond, E., McKinnon, E., & Nolan, D. (2010). Human immunodeficiency virus treatment—induced adipose tissue pathology and lipoatrophy: prevalence and metabolic consequences. *Clinical Infectious Diseases*, *51*(5), 591–599.

Hatzakis, G. E., & Tsoukas, C. M. (2002). Neural networks morbidity and mortality modeling during loss of HIV T-cell homeostasis. *Proceedings of the AMIA Symposium*, 320.

Hejazi, N., & Rajikan, R. (2015). Metabolic Abnormalities in HIV-Infected Populations without or with Antiretroviral Therapy (ART). In *Health of HIV Infected People* (pp. 17–49). Elsevier.

Hirose, H., Takayama, T., Hozawa, S., Hibi, T., & Saito, I. (2011). Prediction of

Universiti Malaya

metabolic syndrome using artificial neural network system based on clinical data including insulin resistance index and serum adiponectin. *Computers in Biology and Medicine*, *41*(11), 1051–1056.

HIV Replication Cycle. (2018). National Institute of Allergy and Infectious Diseases. https://www.niaid.nih.gov/diseases-conditions/hiv-replication-cycle

Hruz, P. W. (2011). Molecular mechanisms for insulin resistance in treated HIV-infection. *Best Practice & Research Clinical Endocrinology & Metabolism*, *25*(3), 459–468.

Hsue, P. Y., Hunt, P. W., Schnell, A., Ho, J., Wu, Y., Hoh, R., Martin, J. N., Deeks, S. G., & Ganz, P. (2010). Inflammation is associated with endothelial dysfunction among individuals with treated and suppressed HIV infection. *Journal of the American College of Cardiology*, *55*(10S), A168-E1577.

Huang, Y.-S., Yang, J.-J., Lee, N.-Y., Chen, G.-J., Ko, W.-C., Sun, H.-Y., & Hung, C.-C. (2017). Treatment of Pneumocystis jirovecii pneumonia in HIV-infected patients: a review. *Expert Review of Anti-Infective Therapy*, *15*(9), 873–892.

Husain, N. E. O. S., & Ahmed, M. H. (2015). Managing dyslipidemia in HIV/AIDS patients: challenges and solutions. *HIV/AIDS (Auckland, NZ)*, *7*, 1.

Jagdeo, J., Ho, D., Lo, A., & Carruthers, A. (2015). A systematic review of filler agents for aesthetic treatment of HIV facial lipoatrophy (FLA). *Journal of the American Academy of Dermatology*, *73*(6), 1040–1054.

Jebb, S. A., Goldberg, G. R., & Elia, M. (1993). DXA measurements of fat and bone mineral density in relation to depth and adiposity. In K. J. Ellis & J. D. Eastman (Eds.), *Human body composition: in vivo methods, models, and assessment* (pp. 115–119).

Universiti Malaya

Springer.

John, M., Nolan, D., & Mallal, S. (2001). Antiretroviral therapy and the lipodystrophy syndrome. *Antiviral Therapy*, *6*(1), 9–20.

Johnson, J. A., Albu, J. B., Engelson, E. S., Fried, S. K., Inada, Y., Ionescu, G., & Kotler, D. P. (2004). Increased systemic and adipose tissue cytokines in patients with HIV-associated lipodystrophy. *American Journal of Physiology Endocrinology and Metabolism*, *286*(2), E261-71. https://doi.org/10.1152/ajpendo.00056.2003

Jothi, N., & Husain, W. (2015). Data mining in healthcare–a review. *Procedia Computer Science*, *72*, 306–313.

Kakuda, T. N. (2000). Pharmacology of nucleoside and nucleotide reverse transcriptase inhibitor-induced mitochondrial toxicity. *Clinical Therapeutics*, *22*(6), 685–708. https://doi.org/10.1016/S0149-2918(00)90004-3

Karsoliya, S. (2012). Approximating number of hidden layer neurons in multiple hidden layer BPNN architecture. *International Journal of Engineering Trends and Technology*, *3*(6), 714–717.

Kaser, A., Brandacher, G., Steurer, W., Kaser, S., Offner, F. A., Zoller, H., Theurl, I., Widder, W., Molnar, C., & Ludwiczek, O. (2001). Interleukin-6 stimulates thrombopoiesis through thrombopoietin: role in inflammatory thrombocytosis. *Blood, The Journal of the American Society of Hematology*, *98*(9), 2720–2725.

Kerr, R., Stirling, D., & Ludlam, C. A. (2001). Interleukin 6 and haemostasis. *British Journal of Haematology*, *115*(1), 3–12.

Khovidhunkit, W., Kim, M.-S., Memon, R. A., Shigenaga, J. K., Moser, A. H., Feingold,

Universiti Malaya

K. R., & Grunfeld, C. (2004). Effects of infection and inflammation on lipid and lipoprotein metabolism: mechanisms and consequences to the host. *The Journal of Lipid Research*, *45*(7), 1169–1196.

Kiebzak, G. M., Leamy, L. J., Pierson, L. M., Nord, R. H., & Zhang, Z. Y. (2000). Measurement precision of body composition variables using the lunar DPX-L densitometer. *Journal of Clinical Densitometry*, *3*(1), 35–41.

Klatzmann, D., Barre-Sinoussi, F., Nugeyre, M. T., Danquet, C., Vilmer, E., Griscelli, C., Brun-Veziret, F., Rouzioux, C., Gluckman, J. C., & Chermann, J.-C. (1984). Selective tropism of lymphadenopathy associated virus (LAV) for helper-inducer T lymphocytes. *Science*, *225*(4657), 59–63.

Koehrsen, W. (2018). Overfitting vs. underfitting: A complete example. *Towards Data Science*.

Koethe, J. R., Dee, K., Bian, A., Shintani, A., Turner, M., Bebawy, S., Sterling, T. R., & Hulgan, T. (2013). Circulating interleukin-6, soluble CD14, and other inflammation biomarker levels differ between obese and nonobese HIV-infected adults on antiretroviral therapy. *AIDS Research and Human Retroviruses*, *29*(7), 1019–1025.

Kornblith, A. E., Addo, N., Dong, R., Rogers, R., Grupp-Phelan, J., Butte, A., Callcut, R., & Arnaout, R. (2020). Development and Validation of a Deep Learning Model for Automated View Classification of Pediatric Focused Assessment with Sonography for Trauma (FAST). *MedRxiv*.

LaForgia, J., Dollman, J., Dale, M. J., Withers, R. T., & Hill, A. M. (2009). Validation of DXA body composition estimates in obese men and women. *Obesity*, *17*(4), 821–826.

Lagathu, C., Eustace, B., Prot, M., Frantz, D., Gu, Y., Bastard, J.-P., Maachi, M.,

Universiti Malaya

Azoulay, S., Briggs, M., & Caron, M. (2007). Some HIV antiretrovirals increase oxidative stress and alter chemokine, cytokine or adiponectin production in human adipocytes and macrophages. *Antiviral Therapy*, *12*(4), 489.

Lake, J. E., & Currier, J. S. (2013). Metabolic disease in HIV infection. *The Lancet Infectious Diseases*, *13*(11), 964–975. https://doi.org/10.1016/S1473-3099(13)70271-8

Landi, A., Piaggi, P., Laurino, M., & Menicucci, D. (2010). Artificial neural networks for nonlinear regression and classification. *2010 10th International Conference on Intelligent Systems Design and Applications*, 115–120.

Larder, B., Wang, D., Revell, A., Montaner, J., Harrigan, R., De Wolf, F., Lange, J., Wegner, S., Ruiz, L., Pérez-Elías, M. J., Emery, S., Gatell, J., Monforte, A. D. A., Torti, C., Zazzi, M., & Lane, C. (2007). The development of artificial neural networks to predict virological response to combination HIV therapy. *Antiviral Therapy*, *12*(1), 15–24. https://doi.org/10.1017/CBO9780511628108.009

Laskey, M. A., Lyttle, K. D., Flaxman, M. E., & Barber, R. W. (1992). The influence of tissue depth and composition on the performance of the Lunar dual-energy X-ray absorptiometer whole-body scanning mode. *European Journal of Clinical Nutrition*, *46*(1), 39–45.

Leclercq, P., Goujard, C., Duracinsky, M., Allaert, F., L'henaff, M., Hellet, M., Meunier, J. P., Carret, S., Thevenon, J., & Ngo Van, P. (2013). High prevalence and impact on the quality of life of facial lipoatrophy and other abnormalities in fat tissue distribution in HIV-infected patients treated with antiretroviral therapy. *AIDS Research and Human Retroviruses*, *29*(5), 761–768.

Lee, G. A., Schwarz, J.-M., Patzek, S., Kim, S., Dyachenko, A., Wen, M., Mulligan, K.,

Universiti Malaya

Schambelan, M., & Grunfeld, C. (2009). The acute effects of HIV protease inhibitors on insulin suppression of glucose production in healthy HIV-negative men. *Journal of Acquired Immune Deficiency Syndromes (1999)*, *52*(2), 246.

Leung, V. L., & Glesby, M. J. (2011). Pathogenesis and treatment of HIV lipohypertrophy. *Current Opinion in Infectious Diseases*, *24*(1), 43.

Li, J., Cheng, J. H., Shi, J. Y., & Huang, F. (2012). Brief introduction of back propagation (BP) neural network algorithm and its improvement. *Advances in Intelligent and Soft Computing*, *169 AISC*(VOL. 2), 553–558. https://doi.org/10.1007/978-3-642-30223-7_87

Li, Z., & Li, Y. (2020). A comparative study on the prediction of the BP artificial neural network model and the ARIMA model in the incidence of AIDS. *BMC Medical Informatics and Decision Making*, *20*(1), 1–13.

Lichtenstein, K. A. (2005). Redefining lipodystrophy syndrome: risks and impact on clinical decision making. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, *39*(4), 395–400.

Lihn, A. S., Richelsen, B., Pedersen, S. B., Haugaard, S. B., Rathje, G. S., Madsbad, S., Andersen, O., Aina, S., Richelsen, B., Pedersen, S. B., Haugaard, S. B., Rathje, G. S., & Madsbad, S. (2003). Increased expression of TNF-α, IL-6, and IL-8 in HALS: implications for reduced adiponectin expression and plasma levels. *American Journal of Physiology-Endocrinology and Metabolism*, *285*(5), E1072–E1080.

Maartens, G., Celum, C., & Lewin, S. R. (2014). HIV infection: epidemiology, pathogenesis, treatment, and prevention. *The Lancet*, *384*(9939), 258–271.

Mallal, S. A., John, M., Moore, C. B., James, I. R., & McKinnon, E. J. (2000).

Universiti Malaya

Contribution of nucleoside analogue reverse transcriptase inhibitors to subcutaneous fat wasting in patients with HIV infection. *Aids*, *14*(10), 1309–1316.

McComsey, G. A., Kitch, D., Sax, P. E., Tebas, P., Tierney, C., Jahed, N. C., Myers, L., Melbourne, K., Ha, B., & Daar, E. S. (2011). Peripheral and central fat changes in subjects randomized to abacavir-lamivudine or tenofovir-emtricitabine with atazanavir-ritonavir or efavirenz: ACTG Study A5224s. *Clinical Infectious Diseases*, *53*(2), 185–196.

McGinn, T., Jervis, R., Wisnivesky, J., Keitz, S., Wyer, P. C., & for the Evidence-based Medicine Teaching Tips Working Group. (2008). Tips for Teachers of Evidence-based Medicine: Clinical Prediction Rules (CPRs) and Estimating Pretest Probability. *Journal of General Internal Medicine*, *23*(8), 1261–1268. https://doi.org/10.1007/s11606-008-0623-z

Medical Development Division. (2011). *Guidelines For The Management of Adult HIV Infection with Antiretroviral Therapy*.

Mhatre, M. S., Siddiqui, F., Dongre, M., & Thakur, P. (2017). A review paper on artificial neural network: a prediction technique. *International Journal of Scientific and Engineering Research (IJSER)*, *8*, 2229–5518.

Míguez, M. J., Lewis, J. E., Bryant, V. E., Rosenberg, R., Burbano, X., Fishman, J., Asthana, D., Duan, R., Madhavan, N., & Malow, R. M. (2010). Low cholesterol? Don't brag yet... hypocholesterolemia blunts HAART effectiveness: a longitudinal study. *Journal of the International AIDS Society*, *13*(1), 1–10.

Milliken, L. A., Going, S. B., & Lohman, T. G. (1996). Effects of variations in regional composition on soft tissue measurements by dual-energy X-ray absorptiometry. *International Journal of Obesity and Related Metabolic Disorders: Journal of the*

Universiti Malaya

International Association for the Study of Obesity, *20*(7), 677–682.

Munawwar, A., & Singh, S. (2016). Human herpesviruses as copathogens of HIV infection, their role in HIV transmission, and disease progression. *Journal of Laboratory Physicians*, *8*(1), 5–18.

Mupere, E., Zalwango, S., Chiunda, A., Okwera, A., Mugerwa, R., & Whalen, C. (2010). Body composition among HIV-seropositive and HIV-seronegative adult patients with pulmonary tuberculosis in Uganda. *Annals of Epidemiology*, *20*(3), 210–216.

Mutimura, E., Anastos, K., Lin, Z., Cohen, M., Binagwaho, A., & Kotler, D. P. (2010). Effect of HIV infection on body composition and fat distribution in Rwandan women. *Journal of the International Association of Physicians in AIDS Care*, *9*(3), 173–178.

Nansseu, J. R., Bigna, J. J., Kaze, A. D., & Noubiap, J. J. (2018). Incidence and risk factors for prediabetes and diabetes mellitus among HIV-infected adults on antiretroviral Therapy. *Epidemiology*, *29*(3), 431–441.

Neye, Y., Düfer, M., Drews, G., & Krippeit-Drews, P. (2006). HIV protease inhibitors: suppression of insulin secretion by inhibition of voltage-dependent K+ currents and anion currents. *Journal of Pharmacology and Experimental Therapeutics*, *316*(1), 106–112.

Nguyen, K. A., Peer, N., Mills, E. J., & Kengne, A. P. (2016). A meta-analysis of the metabolic syndrome prevalence in the global HIV-infected population. *PLoS ONE*, *11*(3), e0150970. https://doi.org/10.1371/journal.pone.0150970

Nsagha, D. S., Assob, J. C. N., Njunda, A. L., Tanue, E. A., Kibu, O. D., Ayima, C. W., & Ngowe, M. N. (2015). Risk factors of cardiovascular diseases in HIV/AIDS patients on HAART. *The Open AIDS Journal*, *9*, 51.

Universiti Malaya

Osuji, F. N., Onyenekwe, C. C., Ahaneku, J. E., & Ukibe, N. R. (2018). The effects of highly active antiretroviral therapy on the serum levels of pro-inflammatory and anti-inflammatory cytokines in HIV infected subjects. *Journal of Biomedical Science*, *25*(1), 1–8.

Paiardini, M., & Müller-Trutwin, M. (2013). HIV-associated chronic immune activation. *Immunological Reviews*, *254*(1), 78–101.

Palella Jr, F. J., Delaney, K. M., Moorman, A. C., Loveless, M. O., Fuhrer, J., Satten, G. A., Aschman, D. J., Holmberg, S. D., & Investigators, H. I. V. O. S. (1998). Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. *New England Journal of Medicine*, *338*(13), 853–860.

Pasini, A. (2015). Artificial neural networks for small dataset analysis. *Journal of Thoracic Disease*, *7*(5), 953.

Patel, P., Borkowf, C. B., Brooks, J. T., Lasry, A., Lansky, A., & Mermin, J. (2014). Estimating per-act HIV transmission risk: a systematic review. *AIDS (London, England)*, *28*(10), 1509.

Pedersen, K. K., Pedersen, M., Trøseid, M., Gaardbo, J. C., Lund, T. T., Thomsen, C., Gerstoft, J., Kvale, D., & Nielsen, S. D. (2013). Microbial translocation in HIV infection is associated with dyslipidemia, insulin resistance, and risk of myocardial infarction. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, *64*(5), 425–433.

Pontes Signorini, D. J. H., Miranda Monteiro, M. C., de Andrade, M. de F. C., Signorini, D. H., & Eyer-Silva, W. de A. (2012). What should we know about metabolic syndrome and lipodystrophy in AIDS? *Revista Da Associação Médica Brasileira*, *58*(1), 70–75. https://doi.org/10.1590/s0104-42302012000100017

Universiti Malaya

Poorolajal, J., Hooshmand, E., Mahjub, H., Esmailnasab, N., & Jenabi, E. (2016). Survival rate of AIDS disease and mortality in HIV-infected patients: a meta-analysis. *Public Health*, *139*, 3–12.

Rajasuriar, R., Chong, M. L., Ahmad Bashah, N. S., Aziz, A., Siti, A., Mcstea, M., Lee, E. C. Y., Wong, P. L., Azwa, I., & Omar, S. (2017). Major health impact of accelerated aging in young HIV-infected individuals on antiretroviral therapy. *Aids*, *31*(10), 1393–1403.

Rani, V., Deep, G., Singh, R. K., Palle, K., & Yadav, U. C. S. (2016). Oxidative stress and metabolic disorders: pathogenesis and therapeutic strategies. *Life Sciences*, *148*, 183–193.

Rivero-Juárez, A., Guijo-Rubio, D., Tellez, F., Palacios, R., Merino, D., Macías, J., Fernández, J. C., Gutiérrez, P. A., Rivero, A., & Hervás-Martínez, C. (2020). Using machine learning methods to determine a typology of patients with HIV-HCV infection to be treated with antivirals. *Plos One*, *15*(1), e0227188.

Robertson, J., Porter, D., Sattar, N., Packard, C. J., Caslake, M., McInnes, I., & McCarey, D. (2017). Interleukin-6 blockade raises LDL via reduced catabolism rather than via increased synthesis: a cytokine-specific mechanism for cholesterol changes in rheumatoid arthritis. *Annals of the Rheumatic Diseases*, *76*(11), 1949–1952.

Ross, A. C., Armentrout, R., O'Riordan, M. A., Storer, N., Rizk, N., Harrill, D., El Bejjani, D., & McComsey, G. A. (2008). Endothelial activation markers are linked to HIV status and are independent of antiretroviral therapy and lipoatrophy. *Journal of Acquired Immune Deficiency Syndromes (1999)*, *49*(5), 499–506. https://doi.org/10.1097/QAI.0b013e318189a794.Endothelial

Universiti Malaya

Ross, A. C., Rizk, N., O'Riordan, M. A., Dogra, V., El-Bejjani, D., Storer, N., Harrill, D., Tungsiripat, M., Adell, J., & McComsey, G. A. (2009). Relationship between Inflammatory Markers, Endothelial Activation Markers, and Carotid Intima-Media Thickness in HIV-Infected Patients Receiving Antiretroviral Therapy. *Clinical Infectious Diseases*, *49*(7), 1119–1127. https://doi.org/10.1086/605578

Rouzioux, C., & Avettand-Fenoël, V. (2018). Total HIV DNA: a global marker of HIV persistence. *Retrovirology*, *15*(1), 1–7.

Saklayen, M. G. (2018). The global epidemic of the metabolic syndrome. *Current Hypertension Reports*, *20*(2), 1–8.

Samaras, K. (2012). The burden of diabetes and hyperlipidemia in treated HIV infection and approaches for cardiometabolic care. *Current HIV/AIDS Reports*, *9*(3), 206–217.

Samaras, K., Gan, S. K., Peake, P. W., Carr, A., & Campbell, L. V. (2009). Proinflammatory markers, insulin sensitivity, and cardiometabolic risk factors in treated HIV infection. *Obesity*, *17*(1), 53–59. https://doi.org/10.1038/oby.2008.500

Sarngadharan, M. G., Devico, A. L., Bruch, L., Schüpbach, J., & Gallo, R. C. (1985). HTLV-III: the etiologic agent of AIDS. *Retroviruses in Human Lymphoma/Leukemia: Proceedings of the 15th International Symposium of the Princess Takamatsu Cancer Research Fund, Tokyo, 1984*, *15*, 301.

Sarvepalli, S. K. (2015). Deep Learning in Neural Networks: The science behind an Artificial Brain. *Liverpool Hope University, Liverpool*.

Selinger, C., & Katze, M. G. (2013). Mathematical models of viral latency. *Current Opinion in Virology*, *3*(4), 402–407.

Universiti Malaya

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.

Sheela, K. G., & Deepa, S. N. (2013). Review on methods to fix number of hidden neurons in neural networks. *Mathematical Problems in Engineering*, *2013*.

Shi, H., Kokoeva, M. V, Inouye, K., Tzameli, I., Yin, H., & Flier, J. S. (2006). TLR4 links innate immunity and fatty acid–induced insulin resistance. *The Journal of Clinical Investigation*, *116*(11), 3015–3025.

Shuck, J., Iorio, M. L., Hung, R., & Davison, S. P. (2013). Autologous fat grafting and injectable dermal fillers for human immunodeficiency virus–associated facial lipodystrophy: a comparison of safety, efficacy, and long-term treatment outcomes. *Plastic and Reconstructive Surgery*, *131*(3), 499–506.

Singh, Y., Narsai, N., & Mars, M. (2013). Applying machine learning to predict patient-specific current CD 4 cell count in order to determine the progression of human immunodeficiency virus (HIV) infection. *African Journal of Biotechnology*, *12*(23).

Singhania, R., & Kotler, D. P. (2011). Lipodystrophy in HIV patients: its challenges and management approaches. *HIV/AIDS (Auckland, NZ)*, *3*, 135.

Smith, D. E., Hudson, J., Martin, A., Freund, J., Griffiths, M. R., Kalnins, S., Law, M., Carr, A., Cooper, D. A., & Investigators, P. D. G. and. (2003). Centralized assessment of dual-energy X-ray absorptiometry (DEXA) in multicenter studies of HIV-associated lipodystrophy. *HIV Clinical Trials*, *4*(1), 45–49.

Sordo, M. (2002). Introduction to Neural Networks in Healthcare. *Knowledge Management for Medical Care*, *73*(2), 17. https://doi.org/10.1007/s12665-014-3245-2

Universiti Malaya

Stapleton, J. H. (2009). *Linear statistical models* (Vol. 719). John Wiley & Sons.

Tan, S., Sim, K. C., & Gales, M. (2015). Improving the interpretability of deep neural networks with stimulated learning. *2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, 617–623.

Tanaka, T., Narazaki, M., & Kishimoto, T. (2014). *Interleukin-6 BT - Encyclopedia of Medical Immunology: Autoimmune Diseases* (I. R. Mackay, N. R. Rose, B. Diamond, & A. Davidson (Eds.); pp. 579–587). Springer New York. https://doi.org/10.1007/978-0-387-84828-0_38

Tanty, R., & Desmukh, T. S. (2015). Application of artificial neural network in hydrology—A review. *Int. J. Eng. Technol. Res*, *4*, 184–188.

Tate, T., Willig, A. L., Willig, J. H., Raper, J. L., Moneyham, L., Kempf, M. C., Saag, M. S., & Mugavero, M. J. (2012). HIV infection and obesity: Where did all the wasting go? *Antiviral Therapy*, *17*(7), 1281–1289. https://doi.org/10.3851/IMP2348

Taylor, S. A., Lee, G. A., Pao, V. Y., Anthonypillai, J., Aweeka, F. T., Schwarz, J.-M., Mulligan, K., Schambelan, M., & Grunfeld, C. (2010). Boosting-Dose Ritonavir Does Not Alter Peripheral Insulin Sensitivity in Healthy, HIVSeronegative Volunteers. *Journal of Acquired Immune Deficiency Syndromes (1999)*, *55*(3), 361.

Teeraananchai, S., Kerr, S. J., Amin, J., Ruxrungtham, K., & Law, M. G. (2017). Life expectancy of HIV-positive people after starting combination antiretroviral therapy: a meta-analysis. *HIV Medicine*, *18*(4), 256–266.

Tenorio, A. R., Chan, E. S., Bosch, R. J., Macatangay, B. J. C., Read, S. W., Yesmin, S., Taiwo, B., Margolis, D. M., Jacobson, J. M., & Landay, A. L. (2014). Rifaximin has a marginal impact on microbial translocation, t-cell activation and inflammation in HIV-

Universiti Malaya

positive immune non-responders to antiretroviral therapy–ACTG A5286. *The Journal of Infectious Diseases*, *211*(5), 780–790.

Thorn, C. F., Lu, Z., & Whitehead, A. S. (2004). Regulation of the Human Acute Phase Serum Amyloid A Genes by Tumour Necrosis Factor-α, Interleukin-6 and Glucocorticoids in Hepatic and Epithelial Cell Lines. *Scandinavian Journal of Immunology*, *59*(2), 152–158.

Tien, P., Cole, S., Williams, C., Li, R., Justman, J., Cohen, M., Young, M., Rubin, N., Augenbraun, M., & Grunfeld, C. (2003). Incidence of Lipoatrophy and Lipohypertrophy in the Women's Interagency HIV Study. *Journal of Acquired Immune Deficiency Syndromes*, *34*(5), 461–466. https://doi.org/10.1097/00126334-200312150-00003

Trikalinos, T. A., Law, M., Carr, A., Definition, C., & Group, S. (2003). HIV lipodystrophy case definition using artificial Objective : A case definition of HIV lipodystrophy has. *Antiviral Research*, *8*, 435–441.

Tripathi, A., Jerrell, J. M., Liese, A. D., Zhang, J., Rizvi, A. A., Albrecht, H., & Duffus, W. A. (2013). Association of Clinical and Therapeutic Factors with Incident Dyslipidemia in a Cohort of Human Immunodeficiency Virus–Infected and Non-Infected Adults: 1994–2011. *Metabolic Syndrome and Related Disorders*, *11*(6), 417–426. https://doi.org/10.1089/met.2013.0017

Ucciferri, C., Falasca, K., Vignale, F., Di Nicola, M., Pizzigallo, E., & Vecchiet, J. (2013). Improved metabolic profile after switch to darunavir/ritonavir in HIV positive patients previously on protease inhibitor therapy. *Journal of Medical Virology*, *85*(5), 755–759.

UNAIDS. (2019). Global HIV and AIDS statistics 2019 Fact sheet. *Global HIV and AIDs Ststistics, World AIDS Day 2019 Fact Sheet*, *1*(June), 1–6.

Universiti Malaya

Vainer, N., Dehlendorff, C., & Johansen, J. S. (2018). Systematic literature review of IL-6 as a biomarker or treatment target in patients with gastric, bile duct, pancreatic and colorectal cancer. *Oncotarget*, *9*(51), 29820.

Vang, Y. S., Chen, Z., & Xie, X. (2018). Deep learning framework for multi-class breast cancer histology image classification. *International Conference Image Analysis and Recognition*, 914–922.

Vergel, N. (2008). Impact of body changes on the quality of life of HIV-positive treatment-experienced patients-an online community-based survey. *ANTIVIRAL THERAPY*, *13*(8), A85–A85.

Villarroya, F., Domingo, P., & Giralt, M. (2010). Drug-induced lipotoxicity: lipodystrophy associated with HIV-1 infection and antiretroviral treatment. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids*, *1801*(3), 392–399.

Vos, A. G., Idris, N. S., Barth, R. E., Klipstein-Grobusch, K., & Grobbee, D. E. (2016). Pro-inflammatory markers in relation to cardiovascular disease in HIV infection. A systematic review. *PloS One*, *11*(1), e0147484.

Wainstein, M. V, Mossmann, M., Araujo, G. N., Gonçalves, S. C., Gravina, G. L., Sangalli, M., Veadrigo, F., Matte, R., Reich, R., & Costa, F. G. (2017). Elevated serum interleukin-6 is predictive of coronary artery disease in intermediate risk overweight patients referred for coronary angiography. *Diabetology & Metabolic Syndrome*, *9*(1), 1–7.

Wang, G., Wei, W., Jiang, J., Ning, C., Chen, H., Huang, J., Liang, B., Zang, N., Liao, Y., & Chen, R. (2019). Application of a long short-term memory neural network: a burgeoning method of deep learning in forecasting HIV incidence in Guangxi, China.

Universiti Malaya

Epidemiology & Infection, *147*.

Wikby, A., Nilsson, B.-O., Forsey, R., Thompson, J., Strindhall, J., Löfgren, S., Ernerudh, J., Pawelec, G., Ferguson, F., & Johansson, B. (2006). The immune risk phenotype is associated with IL-6 in the terminal decline stage: findings from the Swedish NONA immune longitudinal study of very late life functioning. *Mechanisms of Ageing and Development*, *127*(8), 695–704.

Woerdeman, J., & de Ronde, W. (2011). Therapeutic effects of anabolic androgenic steroids on chronic diseases associated with muscle wasting. *Expert Opinion on Investigational Drugs*, *20*(1), 87–97.

Wynn, G. H., Zapor, M. J., Smith, B. H., Wortmann, G., Oesterheld, J. R., Armstrong, S. C., & Cozza, K. L. (2004). Antiretrovirals, Part 1: Overview, History, and Focus on Protease Inhibitors. *Psychosomatics*, *45*(3), 262–270. https://doi.org/10.1176/appi.psy.45.3.262

Xu, H., Barnes, G. T., Yang, Q., Tan, G., Yang, D., Chou, C. J., Sole, J., Nichols, A., Ross, J. S., & Tartaglia, L. A. (2003). Chronic inflammation in fat plays a crucial role in the development of obesity-related insulin resistance. *The Journal of Clinical Investigation*, *112*(12), 1821–1830.

Yanik, E. L., Achenbach, C. J., Gopal, S., Coghill, A. E., Cole, S. R., Eron, J. J., Moore, R. D., Mathews, W. C., Drozd, D. R., & Hamdan, A. (2016). Changes in clinical context for Kaposi's sarcoma and non-Hodgkin lymphoma among people with HIV infection in the United States. *Journal of Clinical Oncology*, *34*(27), 3276.

Yap, S. H., Abdullah, N. K., McStea, M., Takayama, K., Chong, M. L., Crisci, E., Larsson, M., Azwa, I., Kamarulzaman, A., & Leong, K. H. (2017). HIV/Human

Universiti Malaya

herpesvirus co-infections: Impact on tryptophan-kynurenine pathway and immune reconstitution. *PloS One*, *12*(10), e0186000.

Ying, X. (2019). An overview of overfitting and its solutions. *Journal of Physics: Conference Series*, *1168*(2), 22022.

Zhang, G. P. (2006). Avoiding pitfalls in neural network research. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *37*(1), 3–16.

Universiti Malaya

SUPPLEMENTARY

List of Publications and Papers Presented

Type: Article Journal

Title: Analysis of Fat Mass Value, Clinical and Metabolic Data and Interleukin-6 in HIV-Positive Males using Regression Analyses and Artificial Neural Network

Author: Nurul F. Shamsuddin, Mas S. Mokhtar, Reena Rajasuriar, Wan Safwani WKZ, Fatimah Ibrahim, Adeeba Kamarulzaman and Shahrul B. Kamaruzzaman

Journal: Acta Scientiarum Technology

Status: Accepted

Universiti Malaya