Acta Scientiarum

http://periodicos.uem.br/ojs
ISSN on-line: 1807-8664
Doi: 10.4025/actascitechnol.v44i1.57634

Analysis of fat mass value, clinical and metabolic data and interleukin-6 in HIV-positive males using regression analyses and artificial neural network

Nurul Farhah Shamsuddin, Mas Sahidayana Mohktar*, Reena Rajasuriar, Wan Safwani Wan Kamarul Zaman, Fatimah Ibrahim and Adeeba Kamarulzaman

Department of Biomedical Engineering, Centre for Innovation in Medical Engineering, Faculty of Engineering, University of Malaya, 50603, Kuala Lumpur, Malaysia. *Author for correspondence. E-mail: mas_dayana@um.edu.my

ABSTRACT. The purpose of this study is to analyses the relationship between fat mass and inflammation marker, interleukin-6, clinical and metabolic data in 71 human immunodeficiency virus (HIV)-positive male patients using bivariate linear regression analyses and artificial neural network. The data used consisted of measurements collected from HIV male subjects aged 26 to 69 years, with body mass index (BMI) values between 15.47 and 36.98 kg m^{-2} and the fat mass values between 1.00 kg and 16.70 kg. The bivariate linear regression analyses showed that weight, waist-hip ratio, BMI, triglycerides, high-density lipoprotein and HIV viral load value were significant risk factors associated with the body fat mass in male HIV patients. Furthermore, an in-depth non-linear analysis has been performed using artificial neural network (ANN) to predict fat mass by using the significant predictors as input. ANN model with four hidden neurons obtained the highest mean predictive accuracy percentage of 85.26%. The finding of this study is able to help with the evaluation of the fat mass in the male HIV patients that consequently reflects the patients metabolic-related irregularity and immune response. It is also believed that the outcome from the analysis can help future HIV-related study on the prediction of body fat mass in male HIV patients especially in settings where dual energy X-ray absorptiometry assessments, the standard measurement method for fat mass are not available or affordable.

Keywords: HIV; fat mass; bivariate linear regression analyses; artificial neural network (ANN).

Received on February 1, 2021.
Accepted on March 1, 2021.

Introduction

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 (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, Devico, Bruch, Schüpbach, & Gallo, 1984). 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 shown 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).

Alteration of fat in HIV-positive individuals especially fat associated with subcutaneous adipose tissues has largely been linked to increased inflammation caused by the activation of immune cells and cytokines (Cervia et al., 2010). The macrophage number and the measurement of interleukin-6 (IL-6), interleukin-8 (IL-8) and tumor necrosis factor-α (TNF-α) are detected to be high in HIV-positive individuals with lipodystrophy (Saumoy et al., 2008). Risk factors relating to metabolic indicators also have been associated as potential cause in addition to inflammation factor in complicating mechanism of fat redistribution in people living with HIV (Godfrey et al., 2019). Complex relationship between adipose tissue dysfunction, inflammation and immune

function warrants a focus in fat abnormalities assessment, including how it leads to metabolic diseases among people living with HIV. The condition is further worsened with long term infection and prolong consumption of antiretroviral (ART) medication (Villarroya, Domingo, & Giralt, 2010). 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, Fuentes, Vilahur, Badimon, & Palomo, 2013). Hence, increasing recognition of disturbed fat mass (FM) physiology is strongly associated with adipose tissue mass and implicated with IL-6 in HIV patients with abnormal distribution of body fat. Body fat composition constitutes visceral fat and subcutaneous fat requires evaluation of FM to reflect metabolic-related irregularity and immune response in people living with HIV.

Prediction and diagnosis of medical conditions have its own regulations and rules which researchers, physicians and clinicians must obey according to the Clinical Prediction Rules (CPRs). CPRs is a set of rules or procedures concerned with how researchers and physicians apply their judgement in treating their patients based on information obtained which includes medical signs, symptoms and physiological scores (McGinn et al., 2008). These clinical findings are used in diagnoses and prognostics using mathematical tools such as linear regression or similar linear statistical methods. By following the proposed set of rules and guidelines, physicians could advance and thus, prevent the morbidity and mortality of said disease. However, it is reported that physicians encountered lower accuracy and reliability with CPRs in their practise (Adams & Leveson, 2012). The key reason cited for poor accuracy includes the difference between the initial and latter patient population under treatment whereby the CPR is not extensible and adaptive enough to adjust the difference. On the other hand, engineering approach has been considered and studied extensively as an alternative prognostic method in health management field. Among its adaptive mathematical methods used is the Artificial Neural Network (ANN), a data processing model mapped according to brain behavioural system and act as prediction tool with complex pattern model (Santos, Duarte, Faria, & Eduardo, 2009). The main reasons are it can mimic non-linear relationships, handle adaptive learning, pattern recognition and classification, attributes regarded as important in building medical predictive models.

Therefore, the aim of this study is to analyse the relationship between FM with the marker of inflammation, IL-6, metabolic and clinical markers including body-mass index (BMI) and waist-hip ratio (WHR) in HIV-positive males using the bivariate linear regression analyses and ANN.

Material and methods

Subjects

The data used consisted of 71 HIV-positive male subjects aged 26 to 69 years, with BMI values between 15.47 and 36.98 kg m^{-2} and the FM values between 1.00 kg and 16.70 kg with an average value of 11.67 kg. The study was conducted using data acquired through the Malaysian HIV and Aging Study (MHIVA) in collaboration with Biomedical Engineering Department, Engineering Faculty, Universiti Malaya. Details of the study have previously been published (Rajasuriar et al., 2017). Briefly, the study encompassed recruitment of people living with HIV who were on routine follow up at the Infectious Diseases Unit, University of Malaya Medical Centre. All participants fulfilled the following inclusion criteria; age > 25 years, on suppressive antiretroviral therapy 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 dual energy X-ray absorptiometry (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 Ethylenediaminetetraacetic acid (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. The body composition DEXA scan data consists of FM was extracted from iPacs system database of University Malaya Medical Centre. The data were completely anonymized, and informed consent has been obtained from all individuals included in this study. The study complied with the national

regulation and University Malaya policies and was performed in accordance with the tenets of the Helsinki Declaration, and has been approved by the Ethics Committee of University Malaya Research Ethic Board (Approval number: MEC 20151-937).

Data analysis

Descriptive statistics were calculated for the variables involved in this study comprised of FM, IL-6, clinical and metabolic parameters and expressed as means ± standard deviation (SD) as shown in Table 1. Bivariate linear regression analyses were performed to test simple hypotheses of association between FM and clinical, metabolic and inflammation parameters. The independent variables consisted of clinical variables (age, height, weight, WHR, BMI, administered duration of combination ART by months, baseline cluster of differentiation 4 (CD4) cell count, current CD4 cell count, $\log_{10}$ baseline HIV viral load and $\log_{10}$ current HIV viral load) and metabolic variables (glucose, triglycerides, high-density lipoprotein (HDL), low-density lipoprotein (LDL) and total cholesterol). The dependent variable consisted of body composition variable, FM collected from whole-body DEXA scan. All tests were two-tailed and a p value < 0.05 was considered as statistically significant.

Table 1. Subject characteristics comprising clinical, metabolic, body composition and interleukin-6 characteristics for HIV-positive male subjects. 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)	170.20	±	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})	543.10	±	28.73
Log_{10} baseline HIV viral load (copies mL^{-1})	4.50	±	1.52
Log_{10} current HIV viral load (copies mL^{-1})	1.32	±	0.10
Body composition			
FM (kg)	17.28	±	0.89
Metabolic			
Glucose (mmol L^{-1})	5.66	±	0.19
Triglycerides (mmol L^{-1})	1.79	±	0.12
Total cholesterol (mmol L^{-1})	5.08	±	0.11
HDL cholesterol (mmol L^{-1})	1.33	±	0.04
LDL cholesterol (mmol L^{-1})	2.95	±	0.09
Inflammation marker			
IL-6 (ng mL^{-1})	2786.52	±	303.31

ANN modelling of relationship between body fat composition and inflammation involves the interaction of many diverse in nature variables. The ANN model utilized the independent variables that are deem to be significantly associated with FM based on the bivariate linear regression analysis.

The ANN model performance were assessed through the model error function, error sum of squares (SSE), relative error (RE) and mean predictive accuracy percentage (MPA%). The error sum of squares was given by the following equation:

$$E = \sum_n \varepsilon_n = \sum_n (d_n - y_n)^2 \quad (1)$$

where $\in_n$ is output error, d_n is desired output and y_n is predicted output. MPA% was calculated using the following equation:

$$MPA\% = \frac{1}{N}\sum_{i=2}^{N}\left(1 - \frac{|FM_P - FM_D|}{FM_D}\right) 100\% \quad (2)$$

where N was the number of subjects, FM_P was the predicted FM value estimated by ANN, FM_D was desired FM measured by DEXA method. Data was analyzed using software SPSS version 25 (IBM Corporation, Armonk, NY, USA).

Model design and building

ANN models consisting of one input layer, one hidden layer and one output layer topology were constructed with varied different number of hidden neurons in the hidden layer with NH = 1, 2,....10. In this study, an ANN model was developed explore the accuracy of using the seven significant variables associated to FM based on the bivariate linear regression analysis. Feed-forward back-propagation algorithm was developed for the ANN model involving three-layers topology acting as model processor with certain role and function. Each layer consisted of a perceptron produced when several of these neurons interacted with each other in a layer of linear threshold unit. When several layers of perceptron connected in complex interaction involving reinforcement of weights that led to correct behaviour of model during training, a multilayer of perceptron was formed. The network system learned over time how to produce the desired output by using this approach. This multilayer perceptron had several layers which are input layer, output layer and hidden layer. 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. The multilayer perceptron served as a collection of neurons in multiple layers that are connected to each other to form a very complex behaviour due to a lot of different possibilities for all the weights for all the different connections (Popescu, Balas, Perescu-Popescu, & Mastorakis, 2009).

Gradient descent was applied to train the neural network model using chain rule. Gradient descent applied the machine learning optimization technique and it was used to find the optimal set of parameters for a model to solve (Dkhichi & Oukarfi, 2014). It measured the error of the model learning system by plotting a certain function that showed the possible results a neural network model can produce by different set of parameters. By using this method, the model was tuned accordingly to identify the optimal values of parameters that produced the optimal result. 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. New value was then generated at the next pass or epoch of training with new optimized weight and connections of neurons. The training steps were repeated until the system converged and produced output with the least amount of error.

It was crucial for the model to avoid error caused by either overfitting or underfitting. Overfitting occurred when the network model considered the noises or corrupt data and treated it as true data, hence incorrect interpretation of function mapping (Lawrence & Giles, 2000). Underfitting was in opposite with the former with the model underfitted the training data and unable to identify the data trend making the model unfitted for new datasets (Narayan & Tagliarini, 2005). Train, test and holdout was a method to prevent overfitting/underfitting and predicted the measurement of the model's ability to perform on data it never encountered (Moore, 2001). 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 behaviour 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.

However, it was still vulnerable to overfitting due to specific train/test split or the training data was not the representative of the entire data set or the training data set has a skew distribution. 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. K-fold cross validation is a technique that split the data to numbers of data set which the number of the data set is symbolized with K (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 neural network 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. No

definite and exact protocols to be followed when choosing the optimal artificial neural network architecture as long as the basic artificial neural network topology was followed. It was considered a good neural network model if it contained one hidden layer and suitable number of hidden neurons for a regression neural network modelling function (Karsoliya, 2012). Trial-and-error method was used in this study and its existing, wide use was considered an appropriate method to find optimal number of hidden neurons (Sheela & Deepa, 2013). Hyperbolic tangent activation function was used in the hidden and output layer. The neural network function was picked 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. The three partitions corresponded to training dataset, testing dataset and holdout dataset. The number of subjects 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 neural network model (Figrue 1).

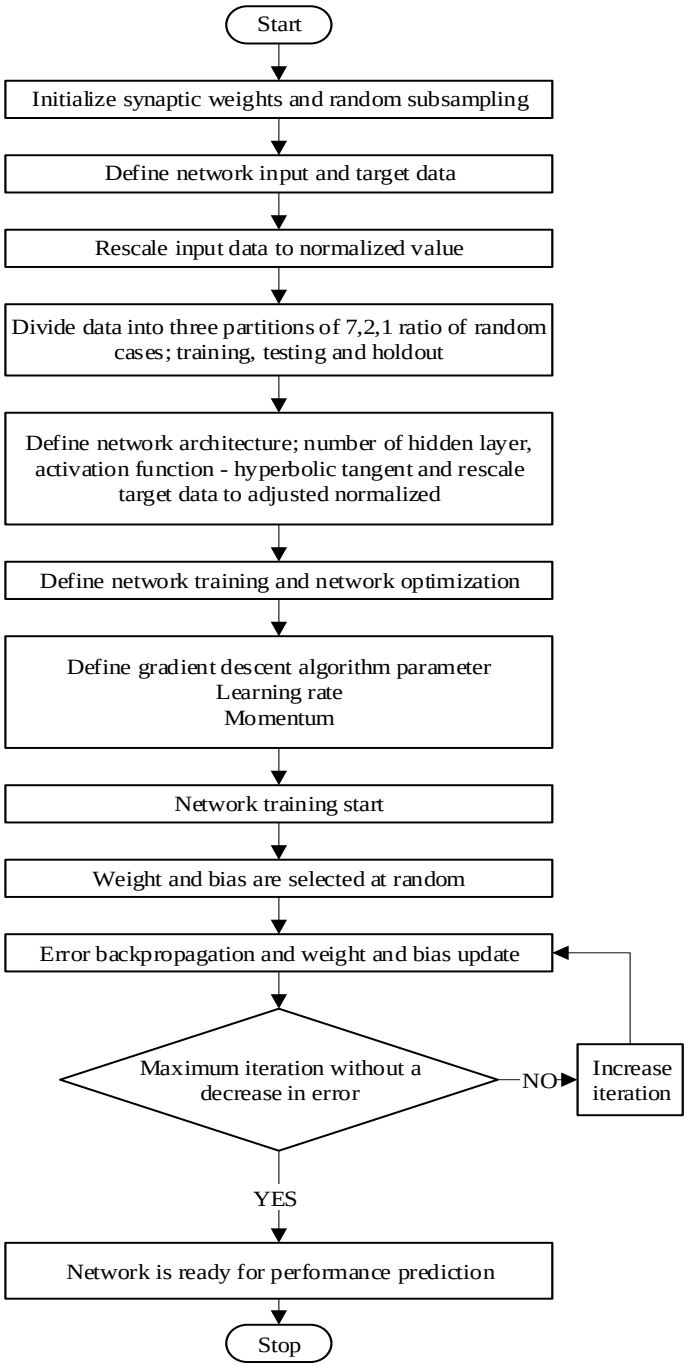

Figure 1. Flowchart of the multi-layer perceptron artificial neural network.

Results

The results show the analysis outcome from the bivariate linear regression and the ANN.

Table 2 displays the bivariate linear regression between FM and the 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.

The total variance, r^2 for FM explained by the study variables has 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 with FM with the rest of the variables – age (0%), height (8.5%), duration on ART (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%). 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 (VIF) where the value is at 1.0 for all the study variables. The performance of the ANN models was identified through SSE, RE and MPA% as shown in Table 3 and Figure 2.

Table 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
FM	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 < .05 DV - dependent variable, IV- independent variable, coef. -coefficient, SE – standard error.

Table 3. Performance of neural networks using significant variables from bivariate linear regression analysis.

Network model	Number of hidden nodes	SSE_{train}	SSE_{test}	$RE_{holdout}$	MPA%
Model 1	1	1.115	0.006	0.483	84.29%
Model 2	2	0.934	0.175	0.235	83.73%
Model 3	3	1.156	0.036	0.563	83.80%
Model 4	4	0.756	0.146	0.371	85.26%
Model 5	5	0.722	0.152	0.380	84.32%
Model 6	6	0.723	0.218	2.660	85.23%
Model 7	7	0.801	0.106	0.159	84.85%
Model 8	8	1.207	0.063	0.440	80.21%
Model 9	9	0.772	0.131	0.179	84.28%
Model 10	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%

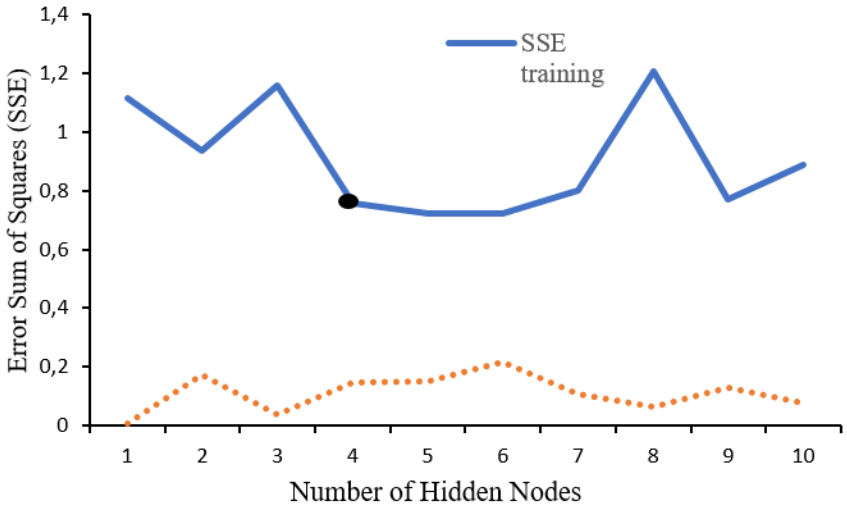

Figure 2. Graph of error sum of squares (SSE) against number of hidden nodes in ANN hidden layer – ANN models using significant variables from bivariate linear regression analysis as factor. Straight line represent errors from training subset and dotted line represents error from testing subset. Best model is represented by a dot.

Overall, 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. The single-hidden layer ANN model with four hidden neurons in the ANN Model obtained the highest MPA% of 85.26% and 0.756 SSE for its training subset, 0.146 SSE for its testing error and 0.371 RE error for its holdout subset.

Discussion

ANN approach is chosen because it tackles the non-linear dependencies which exist in the data where the model proposed fit better than the standard linear statistical model. This study used feed-forward neural network with backpropagation technique to achieve convergence and obtain desired solution by following step-by-step neural modelling algorithm. The results demonstrated that in addition to bivariate linear regression analyses, ANN can be used as a tool to analyze the relationship between the body FM of HIV-positive males with the clinical, metabolic and inflammation components. The analysis using the ANN model are considered important because of its ability to improve the efficiency and efficacy of predicting the body fat composition and manage suitable interventions in HIV-positive males. By predicting the FM, the body fat composition of HIV-positive males can be altered by clinically monitoring the risk factors especially through the significant risk factors associated with the body FM as determined by 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 HIV-positive males. The significant association of these factors suggest the best clinical measure to control in comparison to the other risk factors. Furthermore, employment of ANN prediction reflects faster change and faster accurate assessment of FM value when the risk factors are monitored and controlled in HIV-positive males using timely interventions such as in medication therapy, additional testing, and medical treatment.

The significant anthropometric indicators, weight, WHR and BMI were included to effectively predict FM in HIV-positive males. From the clinical perspective, the parameters are good indicator to assess body fat composition even though the parameters are not able to distinguish the abdominal subcutaneous fat, total abdominal fat and total body fat (Beraldo et al., 2016).

It is known that metabolic parameters – glucose, triglycerides, HDL, LDL and total cholesterol is substantial in body fat composition of HIV-positive males. Even though the pathogenesis of body fat and metabolic changes occurring in HIV-positive individual 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 HIV-positive individuals taking ART treatment (Hejazi & Rajikan, 2015). It is on this account the metabolic factor of triglycerides and HDL are significant in modulating FM and in overall, the body composition development in HIV-positive individuals. Consumption of lipid-lowering agent, for example statins usually is taken alongside ART regimens as management measures for metabolic syndrome in HIV-positive individuals in addition to healthy lifestyle choice (Calza et al., 2016).

The clinical factor, the HIV viral load value 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. Slower progression of HIV disease rate is associated with higher FM value in ART naïve (Martinez et al., 2016) but the consequence of increasing FM and redistribution of fat density on viral load count in people living with HIV on ART has yet been conclusive with several results have been inconsistent. Although CD4 cell count is raised over prolong period of time with increase in BMI among people living with HIV taking ART (Koethe et al., 2016) but the association remains unfavorable and adverse which translates to negative health outcomes, including metabolic-related diseases.

One area required focus was the presence of inflammation marker, IL-6 as one of risk factors in determining overall body fat composition in HIV-positive males. The role of IL-6 is dictated in the body fat composition of HIV-positive males by having inter-relationship and complex interaction starting from the generation of inflammation mediator cytokines which the IL-6 is one of the participants either secreted by adipose tissues or present in body liquid plasma becomes the changes influencing the fat morphology, adipocyte differentiation and lipid profile infecting the person having HIV (Cervia et al., 2010). In this aspect, the IL-6 does not show significant association with the FM in HIV-positive male patients. Other factors such as microbial translocation from the gut, changes in gut microbiome (Tenorio et al., 2015) and co-infection with latent viruses like human herpes viruses (Munawwar & Singh, 2016) should be considered together with the changes in body composition in influencing the IL-6 value in HIV-positive males.

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. Use of systematic experimentation (Sheela & Deepa, 2013) is recommended to discover 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 hidden layer. 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 studying the configuration of neural networks used and by testing the transferability of the model hyperparameters, optimal ANN model algorithm is able to be identified.

The training sample used in this study were relatively small. Due to the nature of study, the data gathering of HIV-positive males was incomplete and the collection of informative data was insufficient making the sample size of HIV-positive males 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, Sim, & Gales, 2015).

Conclusion

In conclusion, analysis has been performed to study the relationship between FM and the marker of inflammation, IL-6, metabolic and clinical markers including BMI and WHR in HIV-positive males. The bivariate linear regression analysis shows that weight, WHR, BMI, triglycerides, HDL and HIV viral load value are significant risk factors associated with the body FM in male HIV patients. Furthermore, an in-depth non-linear analysis has been performed using ANN to predict FM by using the significant predictors as input. A single-hidden layer ANN model with four hidden neurons has obtained the highest MPA% of 85.26%. The finding of this study can help the evaluation of FM that consequently reflect the metabolic-related irregularity and immune response in the male HIV patients. It is also believed that the outcome from the analysis is able to help future HIV-related study on the prediction of body FM in male HIV patients especially in settings where DEXA assessments are not available or unaffordable.

References

Adams, S. T., & Leveson, S. H. (2012). Clinical prediction rules. *BMJ, 344*, d8312. DOI: http://doi.org/10.1136/bmj.d8312

Anglemyer, A., Rutherford, G. W., Easterbrook, P. J., Horvath, T., Vitória, M., Jan, M., & Doherty, M. C. (2014). Early initiation of antiretroviral therapy in HIV-infected adults and adolescents: a systematic review. *AIDS, 28*(2), 105–118. DOI: http://doi.org/10.1097/QAD0000000000000232

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. DOI: http://doi.org/10.1371/journal.pone.0149905

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. *The Journal of Antimicrobial Chemotherapy, 71*(6), 1451–1465. DOI: http://doi.org/10.1093/jac/dkv494

Cervia, J. S., Chantry, C. J., Hughes, M. D., Alvero, C., Meyer 3rd, W. A., Hodge, J., ... 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. DOI: https://doi.org/10.1097/INF.0b013e3181ed9f4c

Dkhichi, F., & Oukarfi, B. (2014). Neural network training by gradient descent algorithms: application on the solar cell. *International Journal of Innovative Research in Science Engineering and Technology, 3*(8), 15696–15702. DOI: https://doi.org/10.15680/IJIRSET.2014.0308084

Fauci, A. S., & Folkers, G. K. (2012). Toward an AIDS-free generation. *JAMA, 308*(4), 343–344. DOI: http://doi.org/10.1001/jama.2012.8142

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), 136584. DOI: http://doi.org/10.1155/2013/136584

Godfrey, C., Bremer, A., Alba, D., Apovian, C., Koethe, J. R., Koliwad, S., ... & Grinspoon, S. (2019). Obesity and fat metabolism in human immunodeficiency virus–infected individuals: Immunopathogenic mechanisms and clinical implications. *The Journal of Infectious Diseases, 220*(3), 420–431. DOI: http://doi.org/10.1093/infdis/jiz118

Hejazi, N., & Rajikan, R. (2015). Metabolic abnormalities in HIV-infected populations without or with antiretroviral therapy (ART). In R. R. Watson (Ed.), *Health of HIV Infected People* (Chap. 2, p. 17–49). Cambridge, MA: Elsevier.

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.

Koehrsen, W. (2018, January 28). Overfitting vs. underfitting: a complete example. *Towards Data Science*. Recovered from https://towardsdatascience.com/overfitting-vs-underfitting-a-complete-example-d05dd7e19765

Koethe, J. R., Jenkins, C. A., Lau, B., Shepherd, B. E., Wester, W., Rebeiro, P. F., ... Moore, R. D. (2016). Higher time-updated body mass index: association with improved CD4+ cell recovery on HIV treatment. *Journal of Acquired Immune Deficiency Syndromes, 73*(2), 197-204. DOI: http://doi.org/10.1097/QAI.0000000000001035

Lawrence, S., & Giles, C. L. (2000). Overfitting and neural networks: conjugate gradient and backpropagation. In *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks* [IJCNN 2000], (p. 114-119). DOI: http://doi.org/10.1109/IJCNN.2000.857823

Martinez, S. S., Campa, A., Bussmann, H., Moyo, S., Makhema, J., Huffman, F. G., ... Baum, M. K. (2016). Effect of BMI and fat mass on HIV disease progression in HIV-infected, antiretroviral treatment-naïve adults in Botswana. *The British Journal of Nutrition, 115*(12), 2114–2121. DOI: http://doi.org/10.1017/S0007114516001409

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. DOI: https://doi.org/10.1007/s11606-008-0623-z

Moore, A. W. (2001). Cross-validation for detecting and preventing overfitting. *School of Computer Science Carneigie Mellon University.* Recovered from https://www.cs.cmu.edu/~./awm/tutorials/overfit10.pdf

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. DOI: http://doi.org/10.4103/0974-2727.176228

Narayan, S., & Tagliarini, G. (2005). An analysis of underfitting in MLP networks. In *Proceedings of the 2005 International Joint Conference on Neural Networks* [IEEE], (p. 984–988). DOI: http://doi.org/10.1109/IJCNN.2005.1555986

Pasini, A. (2015). Artificial neural networks for small dataset analysis. *Journal of Thoracic Disease, 7*(5), 953-960. DOI: http://doi.org/10.3978/j.issn.2072-1439.2015.04.61

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, 28*(10), 1509-1519. DOI: http://doi.org/10.1097/QAD.0000000000000298

Popescu, M.-C., Balas, V. E., Perescu-Popescu, L., & Mastorakis, N. E. (2009). Multilayer perceptron and neural networks. *WSEAS Transactions on Circuits and Systems, 8*(7).

Rajasuriar, R., Chong, M. L., Bashah, N. S. A., Aziz, A., Siti, A., Mcstea, M., … Kamarulzaman, A. (2017). Major health impact of accelerated aging in young HIV-infected individuals on antiretroviral therapy. *AIDS, 31*(10), 1393–1403. DOI: http://doi.org/10.1097/QAD.0000000000001475

Santos, F. L., Duarte, M. L. M., Faria, M. T. C. d., & Eduardo, A. C. (2009). Balancing of a rigid rotor using artificial neural network to predict the correction masses. *Acta Scientiarum. Technology, 31*(2), 151–157. DOI: http://doi.org/10.4025/actascitechnol.v31i2.3912

Sarngadharan, M. G., Devico, A. L., Bruch, L., Schüpbach, J., & Gallo, R. C. (1984). HTLV-III: the etiologic agent of AIDS. In *Proceedings of the Princess Takamatsu Symposia*, (p. 301-308) Tokyo, JP. Retrieved from https://europepmc.org/article/med/6100648

Saumoy, M., López-Dupla, M., Veloso, S., Alonso-Villaverde, C., Domingo, P., Broch, M., … Vidal, F. (2008). The IL-6 system in HIV-1-infection and in HAART-related fat redistribution syndromes. *AIDS, 22*(7), 893-896. DOI: http://doi.org/10.1097/QAD.0b013e3282f4dde7

Sheela, K. G., & Deepa, S. N. (2013). Review on methods to fix number of hidden neurons in neural networks. *Mathematical Problems in Engineering, 2013*(425740). DOI: http://doi.org/10.1155/2013/425740

Tan, S., Sim, K. C., & Gales, M. (2015). Improving the interpretability of deep neural networks with stimulated learning. In *Proceedings of the 2015 IEEE Workshop on Automatic Speech Recognition and Understanding* [ASRU], (p. 617–623). DOI: http://doi.org/10.1109/ASRU.2015.7404853

Tenorio, A. R., Chan, E. S., Bosch, R. J., Macatangay, B. J. C., Read, S. W., Yesmin, S., … Wilson, C. C. (2015). Rifaximin has a marginal impact on microbial translocation, t-cell activation and inflammation in HIV-positive immune non-responders to antiretroviral therapy–ACTG A5286. *The Journal of Infectious Diseases, 211*(5), 780–790. DOI: http://doi.org/10.1093/infdis/jiu515

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. DOI: http://doi.org/10.1016/j.bbalip.2009.09.018

Yap, S. H., Abdullah, N. K., McStea, M., Takayama, K., Chong, M. L., Crisci, E., … Rajasuriar, R. (2017). HIV/Human herpesvirus co-infections: impact on tryptophan-kynurenine pathway and immune reconstitution. *PloS One, 12*(10), e0186000. DOI: http://doi.org/10.1371/journal.pone.0186000

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. DOI: http://doi.org/10.1109/TSMCC.2006.876059