NON-INVASIVE APPROACH TO PREDICT THE CHOLESTEROL LEVEL IN BLOOD USING BIOIMPEDANCE AND NEURAL NETWORK TECHNIQUES

M. S. Mohktar*,‡, F. Ibrahim* and N. A. Ismail†

*Medical Informatics and Biological Micro-Electro-Mechanical Systems Specialized Laboratory, Department of Biomedical Engineering, Faculty of Engineering, University of Malaya, 50603 Kuala Lumpur, Malaysia
†Department of Applied Statistics, Faculty of Economics and Administration, University of Malaya, 50603 Kuala Lumpur, Malaysia

Accepted 21 November 2012
Published 9 December 2013

ABSTRACT
This paper presents a new non-invasive approach to predict the status of high total cholesterol (TC) level in blood using bioimpedance and the artificial neural network (ANN) techniques. The input parameters for the ANN model are acquired from a non-invasive bioelectrical impedance analysis (BIA) measurement technique. The measurement data were obtained from 260 volunteered participants. A total of 190 subject’s data were used for the ANN training purpose and the remaining 70 subject’s data were used for model testing. Six parameters from the BIA parameters were found to be significant predictors for TC level in blood using logistic regression analysis. The six input predictors for the ANN modeling are age, body mass index (BMI), body capacitance, basal metabolic rate, extracellular mass and lean body mass. Four ANN techniques such as the gradient descent with momentum, the resilient, the scaled conjugate gradient and the Levenberg–Marquardt were used and compared for predicting the high TC level in the blood. The finding showed that the resilient method was the best model with prediction accuracy, sensitivity, specificity and area under the curve value obtained from the test data were 82.9%, 85.4%, 79.3% and 0.83%, respectively.

Keywords: Non-invasive; total cholesterol; bioelectrical impedance; artificial neural network; logistic regression.

INTRODUCTION
Cholesterol is an important substance produced by the liver. This substance is used for supporting the body’s cellular functions, production of hormones and also plays a central role for many biochemical processes.

Total cholesterol (TC) is the measure of the amount of cholesterol in the blood, which consists of high density lipoprotein (HDL), low density lipoprotein (LDL), and triglycerides.

According to the Third Report of the National Cholesterol Education Program (NCEP) Expert Panel
on detection, evaluation and treatment of high blood cholesterol in adults (Adult Treatment Panel III) (NCEP ATP III) published in year 2001, the blood TC value is considered as “high” if it exceeds 5.2 mmol/L.2

Excess level of TC can be deposited in the arteries and form a plaque. This plaque narrows the arteries and slows or blocks the blood flow; consequently it decreases the oxygen supply to the heart, brain and other body parts. Furthermore, a heart attack or stroke may occur if the blockage is significant.5 Furthermore, a normal cholesterol level is particularly vital in type 2 diabetic patients who suffer from a condition called diabetic dyslipidemia. The elevation of cholesterol level could increase their risk of obtaining cardiovascular disease.6 Thus it is essential for patients with diabetics to proactively monitor their cholesterol levels.

Currently, the standard TC value measurement uses an invasive technique of withdrawing blood from the subjects and sending it for laboratory analysis. Generally 6 mL of serum is obtained from 15 mL of blood from the subject’s vein. The serum is then processed using a chemical method to obtain the blood TC value.7 This procedure is invasive and the subjects that undergo this procedure may be at risk of developing a bruise or hematoma. Furthermore, the procedure is not suitable for person with diabetes.8 Thus a non-invasive technique which can help determine the status of TC level in blood should be explored and introduced.

One of the potential non-invasive methods available is the bioelectrical impedance analysis (BIA). It is a technique that measure human body electrical conductivity value; which are the essential values used to derive the body composition parameters.9,10 The BIA technique has been widely used in human clinical applications, disease states and in on-going research. Most studies have reported that the technique is a highly promising non-invasive, useful and valid means of monitoring health status in patients with communicable and non-communicable disease.11–14 However, none of the studies explored the ability of BIA parameters in predicting TC level in blood.

Artificial neural networks (ANN) are inspired by the structure and behavior of neurons and are a useful pattern recognition tools. Like the brain, ANN can recognize patterns, manage data and learn. One type of ANN is the multilayer feedforward networks (MFNN) that entails the function of a common and more complex nervous system.15 As reported by Fogel et al., MFNN has outperformed other machine learning methods in the detection of breast cancer,16 and Wei et al. has indicated that MFNN is the most common ANN model used in clinical related research.17 Thus, this paper presents a non-invasive, rapid and safe system to predict the status of high TC level in blood (of more than 5.2 mmol/L) using the non-invasive BIA and ANN modeling techniques.

**MEHODOLOGY**

**Subject Recruitments**

Subjects were recruited through advertisements in the internal University Malaya’s email (uminfo@list.um.edu.my) at University of Malaya, Kuala Lumpur, Malaysia. Subjects who met any of the following criteria were not eligible for the study; pregnant woman, having artificial joints, pins, plates or other types of metal objects in the body, pacemaker or automatic implantable cardiac defibrillator, coronary stent or metal suture material in the heart, have a history of metabolic diseases, or illness such as diarrhea. The recruitment process was done on a volunteering basis from the 1st of January 2006 to the 31st December 2006. A total of 260 subjects (equivalent to 1% of the total University of Malaya population) were then recruited. Permission and consent forms were obtained from each subject before measurement was conducted in the University of Malaya Student Health Clinic (UMSHC).

**Data Collection**

Data collections were conducted at the UMSHC. Subjects were asked to fast and abstain from performing any physical activities in the 9–12 h prior to attendance at the UMSHC. Three types of measurements were performed: (i) anthropometric measurements; (ii) BIA measurements; and (iii) full blood cholesterol investigation testing.

**Anthropometric measurements**

Anthropometric measurements include measurement of height, weight and calculation of body mass index (BMI). BMI was calculated by dividing the subject’s weight in kilograms (kg) by height in meters squared (BMI = kg/m²). Weight was measured to the nearest 0.1 kg, while height was measured to the nearest 0.1 cm and was converted to meters (m). These measurements were taken with the subjects in light clothing and without shoes.

**Bioimpedance measurements**

BIA measurement was performed using the Biodynamics Model 450 bioimpedance analyzer. This measurement
was taken while the subject lies in a supine position on a bed. Two sensor pads and sensor cables (red and black) were placed on the subject’s right hand; one at the joint that connects the finger to the hand (red) and another one slightly at the prominent wrist bone (black). Two more sensor pads were placed on the right foot, one 2.5 cm above the toe line (black) and another slightly above the ankle joint (red). Then, a constant current signal less than 1 mA at a single frequency of 50 kHz was applied to the black sensor cables while the other two red sensor cables were used to detect the signal. This measurement procedure took around 3 min to obtain the results.

The results provide measurement values of: (i) bioelectrical tissue conductivity (resistance (R), reactance (X), body capacitance (BC) and phase angle (PA)); (ii) body composition mass distribution (body cell mass (BCM), extracellular mass (ECM), lean body mass (LBM), fat mass (FM) and basal metabolic rate (BMR)); and (iii) water compartments (intracellular water (IW), extracellular water (EW), total body water (TBW), TBW/lean body mass (TBW/LBM) and TBW/total weight (TBW/TW)).

**Blood cholesterol measurements**

A 15 mL of blood was withdrawn from the subject by a medical practitioner at the Clinical Diagnostic Laboratory, University of Malaya Medical Center. The sample were then sent to a laboratory technician for the analysis of TC levels. The samples were tested for TC using an in vitro quantitative measurement in serum and plasma (Cholesterol Flex® reagent cartridge and Dimension® Clinical Chemistry System machine (Dade Behring) using the principle described by Stadtman). \(^7\)

**Blood TC Reference Value**

The conventional medical reference value for the TC level in blood was used in this study. The “normal” category was assigned if subject’s blood TC value is equal or less than 5.2 mmol/L and the “high” blood TC level is assigned to subject with blood TC value of more than 5.2 mmol/L. \(^2,4\) The numbers 0 and 1 represent the normal and abnormal or high status of blood TC, respectively.

**Data Analysis**

Of the 260 subjects, the first 190 collected subject’s data were used for training purposes and the remaining 70 subject’s data were used for independent testing process.

**Input predictor parameters selection**

Predictors were selected using simple logistic regression method. Independent variables with p-value less or equal to 0.25 were considered for further analysis. \(^18\) Next, all continuous predictors were tested for correlation with each other using simple linear correlation test. One of two correlated predictor (with correlation coefficient (r) equal or more than 0.8) was eliminated to avoid multicollinearities. \(^19\)

The next subsequent analysis was the linearity assumption test. This test was performed on each selected continuous predictor. Any nonlinear predictor was categorized (p-value less of equal to 0.05). \(^20\)

**Artificial neural network**

The ANN employed was MFNN. The model consists of three layers: one input layer, one hidden layer and one output layer. The input of the network is formed by the vector \(\mathbf{x} = (x_1, \ldots, x_i)\); i is the number of selected predictors (continuous predictors were normalized prior of the training of the MFNN). \(^21\) The MFNN model is defined by Eq. (1) and the function in the first and the second layer are defined as Eqs. (2) and (3), respectively. \(^15,22,23\)

\[
TC = f(2) \left[ \sum_{j=1}^{H} w_{kj} f(1) \left( \sum_{i=1}^{j} w_{ji} x_i + w_{j0} \right) + w_{k0} \right],
\]

where

\(j = \text{output in the hidden neuron starting from 1 to } H,\)

\(k_0 = \text{initial output},\)

\(k = \text{number of output},\)

\(H = \text{number of neurons in the hidden layer},\)

\(w = \text{weights of the neural network},\)

\((1) = \text{first layer}, (2) = \text{second layer}.\)

\[
f(1) = \frac{2}{(1 + e^{(-2X-1)})}, \quad (2)
\]

\[
f(2) = \frac{1}{(1 + e^{(-X)})}, \quad (3)
\]

where \(X = \text{input}.\)

An initial cut-off value of 0.5 was used. If the ANN output value was less than 0.5, the case was classified as 0 (normal) otherwise it is classified as 1 (high).

The number of neurons in the hidden layer \((H)\) was optimized using a 10-fold cross-validation technique.
The training data were randomly divided into 10 subgroups containing 19 sets of subject’s data. One group was selected as the test set and the remaining nine groups were used as the training set. The training process is repeated 10 times and each subgroup is used once as a test set. The optimum number of hidden layer is accomplished by computing the highest average of accuracy and the lowest mean squared error (MSE) from the 10 cross-validation loops.

Four techniques were considered in the training process; the gradient descent with momentum method (GDMBP), the resilient method (RBP), the scaled conjugate gradient method (SCGBP) and the Levenberg–Marquardt method (LMBP). The cross-validation process mentioned above was performed for the each of the four different algorithms.

The performances of the ANN models were evaluated by comparing the ANN prediction output with the reference value. Then, the optimized ANN models were tested using the validation data (data collected from the remaining 70 subjects). Furthermore, to quantify the models, receiver operating curves (ROC) technique were used to find the area under the curve (AUROC) value and the standard error (SE) value for each of the ANN models. Then, the best cut off point for each of the ANN model which gave the best accuracy, sensitivity and specificity was selected.

RESULT

From the 260 recruited subjects, 112 were male and 148 were female and 49.2% (n = 128) subjects were having TC more than 5.2 mmol/L and the remaining 50.8% (n = 132) were having TC less than 5.2 mmol/L. These data were divided into training dataset and testing dataset.

Fourteen variables were analyzed. The independent variables consist of the anthropometrics parameters and the BIA parameters. Table 1 lists all the variables involved and shows the mean and the standard deviation values for the training and testing dataset.

The simple logistic regression test result (Table 2) indicates that age, weight, BMI and all of the analyzed BIA parameters produced significant effects (p < 0.25) on TC level in blood, however gender and variables were not predictors to TC level in blood at (p > 0.25).

Correlation tests that were performed on the remaining 12 predictors showed significant correlations between weight and BMI (r = 0.9), weight and BMR (r = 0.9), BMR and RES (r = -0.8), BCM and LBM (r = 0.8), BCM and FM (r = -0.9), ECM and FM (r = -0.8) and LB and FM (r = -0.9). Six out of these 12 predictors were selected based on the evidence obtained from literatures. The selected predictors were age, BMI, BC, BMR, ECM and LBM.

The linearity assumption test supports treating Age, BC, BMR, ECM and LBM as continuous as the linearity assumption is met, though, BMI was categorized (Table 3).

Table 4 shows the accuracy and the MSE obtained from all four training algorithm. The selected number of neuron in the hidden layer for LMBP was 19, both RBP and SCGBP was 18 and GDMBP was 16. The highest accuracy was calculated from the SCGBP and the lowest MSE was obtained from the GDMBP learning algorithm.

### Table 1. Characteristics of the Independent Variables.

<table>
<thead>
<tr>
<th>Variable (Unit)</th>
<th>Training</th>
<th>Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normal TC Level (n = 91)</td>
<td>High TC Level (n = 99)</td>
</tr>
<tr>
<td>Age (Years)</td>
<td>34.6 ± 11.2</td>
<td>41.9 ± 9.4</td>
</tr>
<tr>
<td>Gender</td>
<td>38(Male)</td>
<td>49(Male)</td>
</tr>
<tr>
<td>Height (m)</td>
<td>160.0 ± 8.4</td>
<td>160.6 ± 8.4</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>64.5 ± 14.6</td>
<td>72.5 ± 14.7</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>25.3 ± 6.1</td>
<td>28.1 ± 5.2</td>
</tr>
<tr>
<td>R (Ohms)</td>
<td>564 ± 89</td>
<td>538 ± 95</td>
</tr>
<tr>
<td>Xc (Ohms)</td>
<td>61.6 ± 12.5</td>
<td>60.5 ± 11.5</td>
</tr>
<tr>
<td>BC (pF)</td>
<td>638 ± 161</td>
<td>703 ± 194</td>
</tr>
<tr>
<td>PA (%)</td>
<td>6.4 ± 1.1</td>
<td>6.6 ± 1.1</td>
</tr>
<tr>
<td>BCM (%)</td>
<td>35.1 ± 6.1</td>
<td>33.5 ± 5.1</td>
</tr>
<tr>
<td>EM (%)</td>
<td>38.5 ± 4.6</td>
<td>36.3 ± 4.1</td>
</tr>
<tr>
<td>LBM (%)</td>
<td>73.7 ± 9.5</td>
<td>69.1 ± 8.1</td>
</tr>
<tr>
<td>FM (%)</td>
<td>26 ± 9</td>
<td>30 ± 7</td>
</tr>
<tr>
<td>BMR (cals)</td>
<td>1456 ± 273</td>
<td>1571 ± 320</td>
</tr>
</tbody>
</table>
The testing dataset consists of 29 subjects (41.4%) with high TC level in blood and 41 subjects (58.6%) with normal TC level in blood. Table 5 shows the classification results for the ANN models classified at the 0.5 cut-off prediction. RBP and GDMBP algorithms give sensitivity value of 82.9 and 75.9%, respectively while LMBP and SCGBP shows sensitivity of 65.5%. This indicates that RBP and GDMBP were better in detecting subjects with high TC level in blood.

The highest accuracy, sensitivity and specificity values (at the 0.5 cut-off) are obtained from RBP algorithm; 81.4, 82.9 and 79.3, respectively.

Next, ROC curves were plotted to find the AUROC, the SE and the best cut-off prediction point for each of the ANN model. Table 6 shows the model performances at the selected cut-off point. AUROC for RBP is 0.83, with an optimal cut-off point of 0.55. At this cut-off point, the highest sensitivity (85.4%) and the highest specificity (79.3%) values for the RBP algorithm were obtained.

DISCUSSIONS

The ANN models in this paper were constructed using a hybrid logistic regression and the ANN technique, which is a similar approach performed by Spackman et al.33 The simple logistic regression was used for the pre-selection of variables. Then the significant predictors were considered for the development of the ANN models.

For benchmarking purposes, the performances of the RBP ANN model was compared with a conventional modeling method, where the six selected predictors were re-trained and re-tested using the multiple logistic regression (MLR).34

The accuracy, sensitivity and specificity were calculated by comparing the ANN outputs with the reference standards from the validation data set. The specificity of the MLR model is 97.6%; meanwhile the sensitivity is only 34.5%. Hence, the total accuracy is 71.4%.

Figure 1 shows the comparisons of the classification performance between the RBP ANN model and the MLR model.
Even though the structure of MLR model has similarities to the structure of ANN models, the ANN model produces a higher accuracy and sensitivity as compared to the MLR. This is due to the \( H \) nodes in the ANN model that act as data feature extractors that allows the network to learn implicitly any arbitrarily complex nonlinear relationship between the dependent and predictor variables. The training of the ANN model require much more computing time than the conventional MLR statistical methods and the MLR model has a higher specificity, however, the ANN model is considered more efficient. Considering, in a medical data mining, it is important to have an acceptable sensitivity. This is to avoid the situation that carries the most serious clinical risk by incorrectly predicting subjects as having normal TC level when they are having high TC level in blood that might require clinical intervention.

Nevertheless, the introduction of this non-invasive method to predict blood TC level without the need to collect a blood sample could benefit patient’s point-of-care treatment. The advancement of BIA technology has allow the availability of this device at the patient’s home, consequently it permits patients to perform self-monitoring (could help patients to monitor TC level in blood by means of frequent measurement).

The non-invasive cholesterol monitoring method is of upmost importance for patients with diabetic dyslipidemia who need to closely monitor their cholesterol level but at the same time are not recommended to perform invasive procedure. Consequently, due to frequent monitoring of their cholesterol level, they can manage their disease by restricting their calorie intakes whenever a high blood cholesterol level is detected.

In the future, the ANN model could be further improved by; including other variables such as ethnic group and waist circumference as the inputs parameters, and increasing the number of data sampling.

CONCLUSIONS

In this study, BIA data were collected from 260 healthy, voluntary participants. The collected data were analyzed using logistic regression to find the significant predictors for predicting the level of TC. BMI, BC, BMR, EM and LBM were the parameters for predicting the level of TC. These variables were trained using four types of ANN learning algorithms in order to predict the TC level into “normal” or “high” blood TC level. RBP was found to be the best ANN model that gave the highest accuracy of 82.9%, sensitivity of 85.4%, specificity of 79.3% and AUROC of 0.83. This performance was obtained at a 0.55 cut-off prediction. To conclude, this study has introduced a non-invasive approach to predict TC blood level using BIA and ANN techniques. These techniques could be implemented on chronic disease patient’s point-of-care and especially to patient with diabetes.

ACKNOWLEDGMENT

This research is supported and funded by the Prime Minister’s Department, Malaysia, through a special Grant Project No. 66-02-03-0061/H-00000-37039.

REFERENCES

8. Thadhani RI, Camargo CA, Xavier RJ, Fang LST, Bazari H, Atheroembolic renal failure after invasive procedures:

![Fig. 1](image-url)


