Original article

A predictive hypertension model for patients with dyslipidemia and type 2 diabetes mellitus: a robust hybrid methodology

Wan Muhamad Amir W Ahmad[1], Mohamad Nasarudin Adnan[2],Nuzlinda Abdul Rahman[3], Farah Muna Mohamad Ghazali[4], Nor Azlida Aleng[5], Zainab Mat Yudin @ Badrin[6], Mohammad Khursheed Alam[7,8,9]*

Abstract:

Background:Hypertension is a public health problem used to describe high blood pressure where the blood vessels are persistently increased in force. According to WHO, hypertension has been reported in one in four men and one in five women. Worldwide, hypertension is a common health problem that affects 20-30% of the adult population and more than 5-8% of pregnancies, and it is frequently curable when detected and treated early enough. ***Objective:*** This paper aims to validate the factor associatedwith hypertension status among patients with dyslipidemia and type 2 diabetes mellitus. This could help to improve the prediction of the probability of hypertension among studied patients. ***Material and Methods*:** 39 patients were recruited from the Hospital Universiti Sains Malaysia (USM). In this retrospective study, advanced computational statistical modeling methodologies were used to evaluate data descriptions of several variables such as hypertension, marital status, smoking status, systolic blood pressure, fasting blood glucose, total cholesterol, high-density lipoprotein, alanine transferase, alkaline phosphatase, and urea reading. The R-Studio software and syntax were used to implement and test the hazard ratio. The statistics for each sample were calculated using a combination model that included bootstrap and multiple logistic regression methods. ***Results:*** The statistical strategy showed R demonstrates that regression modeling outperforms an R-squared. It revealed that the hybrid model technique better predicts the outcome when data is partitioned into a training and testing dataset. The variable validation was determined using the well-established bootstrap-integrated MLRtechnique. In this case, eight variables are considered: marital status, systolic blood pressure, fasting blood glucose, total cholesterol, high-density lipoprotein, alanine transferase, alkaline phosphatase, and urea reading. It's important to note that six things affect the hazard ratio: Marital status (β_1: 1.183519; $p< 0.25$), systolic blood pressure (:-0.144516; $p< 0.25$), total cholesterol (β_2: 0.9585890; $p< 0.25$), high-density lipoprotein (:-5.927411; $p< 0.25$), alkaline phosphatase (:-0.008973; $p> 0.25$), and urea reading (:0.064169; $p< 0.25$).There is a 0.003469102 MSE for the linear model in this scenario. ***Conclusion:*** In this study, a hybrid approach combining bootstrapping and multiple logistic regression will be developed and extensively tested. The R syntax for this methodology was designed to ensure that the researcher completely understood the illustration. In this case, a hybrid model demonstrates how this critical conclusion enables us to understand better the utility and relative contribution of the hybrid method to the outcome. The statistical technique used in this study, R, demonstrates that regression modelingoutperforms R-squared values of 0.9014 and 0.00882 for the Predicted Mean Squared Error, respectively. Thus, the study's conclusion establishes the superiority of the hybrid model technique used in the study.

Keywords: Hypertension, Dyslipidemia;Multiple logistic regression;R square, Predicted Mean Square Error; type 2 diabetes mellitus

Bangladesh Journal of Medical Science Vol. 22 No. 02 April'23 Page : 422-431
DOI: https://doi.org/10.3329/bjms.v22i2.65007

Introduction:

High blood pressure or hypertension is a well-known health issue worldwide, and it continues to pose a significant global health burden. Obesity and diabetes increase the risk of cardiovascular diseases, such as stroke and ischemic heart disease.[1,2]. Hypertension-related complications caused more than 9 million deaths, according to studies on the clustering of cardiovascular disease burden in 21 regions. Around 40% of adults over the age of 25 have been clinically diagnosed with hypertension[3,4]. 7.4 million deaths from coronary heart disease and 6.7 million deaths from stroke were also caused by this epidemic. [4,5]. The excellent work of the N.C.D.

Correspondence: Wan Muhamad Amir W Ahmad, School of Dental Sciences, Health Campus, Universiti Sains Malaysia, 16150 Kubang Kerian, Kota Bharu, Kelantan, Malaysia Email: wmamir@usm.my

Risk Factor Collaboration on global trends in blood pressure from 1975 to 2015, which included 1,479 population-based studies, discovered that the number of adults with hypertension increased from 594 million in 1975 to 1.13 billion in 2015 [6]. Furthermore, according to WHO, in 2000, nearly one billion of the world's adult population had hypertension, and almost 1.56 billion people (29.2% of the world's population) are projected to have hypertension by 2025 [7,14]. In the general hypertensive population, hypertension, particularly uncontrolled and untreated hypertension, is associated with an increased risk of total and cardiovascular death[8]. Even though assessing hypertension prevention and controlhas been reported as an important public health issue, finding a suitable strategy for controlling it has been a global problem[21].Furthermore, most of the treatment is only happening in the end stage of hypertension,making the treatment less effective[17]. Large-scale research on the effect of modifiable risk factors for myocardial infarction (52 countries) showed that people with a myocardial infarction are 2.5 times more likely than those with normal blood pressure, regardless of ethnicity, sex, and smoking status with severely raised blood pressure [9].
Prehypertension patients were 1.5 times more likely than those with normal blood pressure to develop cardiovascular disease.[10,11]. A 34-year follow-up of the Framingham Heart Study cohort showed that those in the higher blood pressure quintile had more than twice the risk of congestive heart failure than those in the lower quintile[12]. In 2000, over 180 million Chinese people had hypertension, and by 2025, this number was expected to rise by another 100 million. In 2015, Asia was home to more than 40% of the 1.13 billion adults with hypertension, 226 million living in China alone[6]. Despite the government's policies and hypertension task forces, hypertension in Malaysia has remained high in recent decades [13–15]. Current population-based research in Malaysia shows that hypertension is relatively higher in men, older groups, and those with low household incomes[16]. Other risk factors like obesity, smoking, alcohol consumption, hypercholesterolemia, and diabetes [22]. However, they are not widely investigated how many levels of exposure may replicate the incidence rate of an outcome of interest [19]. On the other hand, previous studies have concentrated exclusively on clinical hypertension in adults, with limited documentation of factors associated with the severity of hypertension, particularly among the vulnerable prehypertensive group. However, previous research has concentrated on clinical hypertension in adults, with scant documentation of factors associated with hypertension's severity and, most importantly, in the prehypertensive group, which is particularly vulnerable.

The term 'dyslipidemia', which has replaced the older term 'hyperlipidemia', describes the abnormal changes in body composition, mainly body fat and lipid profiles. Dyslipidemia is a significant risk factor for coronary heart disease with diabetes

1. Wan Muhamad Amir W Ahmad,School of Dental Sciences, Health Campus, Universiti Sains Malaysia (USM),16150 Kubang Kerian, Kota Bharu, Kelantan, Malaysia. wmamir@usm.my; nasarudinadnan@student.usm.my; farahmuna@student.usm.my; drzainab@usm.my
2. Mohamad Nasarudin Adnan,School of Dental Sciences, Health Campus, Universiti Sains Malaysia (USM),16150 Kubang Kerian, Kota Bharu, Kelantan, Malaysia. wmamir@usm.my; nasarudinadnan@student.usm.my; farahmuna@student.usm.my; drzainab@usm.my
3. Nuzlinda Abdul Rahman, School of Mathematical Sciences, Universiti Sains Malaysia (USM), 11800 Pulau Pinang, Malaysia. nuzlinda@usm.my
4. Farah Muna Mohamad Ghazali, School of Dental Sciences, Health Campus, Universiti Sains Malaysia (USM),16150 Kubang Kerian, Kota Bharu, Kelantan, Malaysia. wmamir@usm.my; nasarudinadnan@student.usm.my; farahmuna@student.usm.my; drzainab@usm.my
5. Nor AzlidaAleng, Faculty of Ocean Engineering Technology and Informatics, Universiti Malaysia Terengganu (UMT), 21030 Kuala Nerus, Terengganu, Malaysia. azlida_aleng@umt.edu.my
6. Zainab Mat Yudin @ Badrin, School of Dental Sciences, Health Campus, Universiti Sains Malaysia (USM),16150 Kubang Kerian, Kota Bharu, Kelantan, Malaysia. wmamir@usm.my; nasarudinadnan@student.usm.my; farahmuna@student.usm.my; drzainab@usm.my
7. Mohammad Khursheed Alam, College of Dentistry, Jouf University, Sakaka, Kingdom of Saudi Arabia. mkalam@ju.edu.sa; dralam@gmail.com
8. Department of Dental Research Cell, Saveetha Institute of Medical and Technical Science, Saveetha Dental College and Hospitals, 600077, Chennai, India.
9. Department of Public Health, Faculty of Allied Health Science, Daffodil international University, Dhaka-1216, Bangladesh

mellitus, which isrelatedto lower levels of high-density lipoprotein (HDL) cholesterol, higher levels of plasma triglycerides, and higher levels of small dense particles of low-density lipoprotein (LDL) cholesterol[18]. Lipid abnormalities are common in diabetes mellitus because insulin resistance or deficiency affects critical enzymes and pathways involved in lipid metabolism[20]. Additionally, it has been proposed that diabetic dyslipidemia's lipid particle composition is more atherogenic than other types of dyslipidemia. This means that even normal lipid levels may be more atherogenic in diabetic individuals than in nondiabetic individuals. It is well established that atherosclerosis and dyslipidemia are causally related. The associated hyperglycemia, obesity, and insulin change all significantly accelerate the progression of atherosclerosis in diabetes[22].

Materials and Methods:

Data Collection

This study looked at data from patients at the Hospital Universiti Sains Malaysia (USM) ambulatory clinic. The trial enrolled 39 people. Table 1 summarises the research variables' data descriptions.

Table 1. Data Description of the selected blood profile

Variable	Code	Description
Hyper	Y	Hypertension 0 = Yes 1 = No
Marital	X_1	Marital status 0=Married 1= Single 2 = Widow/widower
Smoke	X_2	Smoking status 0 = Smoker 1 = Exsmoker 2 = Nonsmoker
Sysbp	X_3	Systolic blood pressure
Fbs	X_4	Fasting blood sugar
Tc	X_5	Total cholesterol
Hdl	X_6	High-density lipoprotein
Alt	X_7	Alanine transferase
Alp	X_8	Alkaline phosphatase
Urea	X_9	Urea reading

Study Design

This research uses a retrospective approach and advanced computational statistical modelling methodologies to build a multilayer perceptron with multiple linear regression methodology. This developed methodology was based on the testing and training dataset,MSE-predicted, and the accuracy value of the mean absolute deviance.The study was approved by the Universiti Sains Malaysia Research Ethics and Human Research Committee (USM/JEPeM/16050184).The patient's privacy is protected, as is the patient's medical condition.

Modeling of Computational Biometry

The data were analyzed using binary logistic regression in the R-Studiosoftware, which included the syntax. The advanced technique is a combination model that uses approaches like bootstrap and multiple logistic regression (MLR) when using this method. The data is separated into two distinct groups using this technique, like the training and testing data. The training data will be utilized for the modeling purpose, and the testing data will be used for validation reasons.This section fits a set of logistic regression models to investigate the underlying association between hypertension and the selected explanatory variables. Let us define the following dichotomous dependent variables

Hyper= 0, if not hypertension patient

Hyper =1 is a hypertension patient

Then let us define the following logistic regressionmodels :

$$\sum_{j=1}^{y} y_i = \sum_{i=1}^{n} \pi(x_i)$$

$$\pi(x_i) = \frac{\exp(\beta_0 + \beta_1(Marital) + \beta_2(Sysbp) + \beta_3(Fbs) + \beta_4(Tc) + \beta_5(Hdl) + \beta_6(Alt) + \beta_7(Alp) + \beta_8(Urea))}{1 + \exp(\beta_0 + \beta_1(Marital) + \beta_2(Sysbp) + \beta_3(Fbs) + \beta_4(Tc) + \beta_5(Hdl) + \beta_6(Alt) + \beta_7(Alp) + \beta_8(Urea))}$$

The estimate logit is given by

$$\hat{g}(x) = \beta_0 + \beta_1(Marital) + \beta_2(Sysbp) + \beta_3(Fbs) + \beta_4(Tc) + \beta_5(Hdl) + \beta_6(Alt) + \beta_7(Alp) + \beta_8(Urea) \ldots\ldots(1)$$

The estimated parameter for equation (1) is given in Table 2.

Bootstrap

Once a population sample has been selected at random, bootstrap begins by calculating sample statistics. A pseudo-population is created by using numerous substitution samples, which are then reproduced repeatedly following several replications of the original samples. In the case of replacement sampling, random sampling results in samples that differ from the original sample. The bootstrap calculates statistics for each sample as it is drawn with replacement, and it is used to draw samples with replacement[6,7].The result for the model is

displayed in Table 2. The logistic regression is fitting through the R Software.The complete step by step method is given as follows :

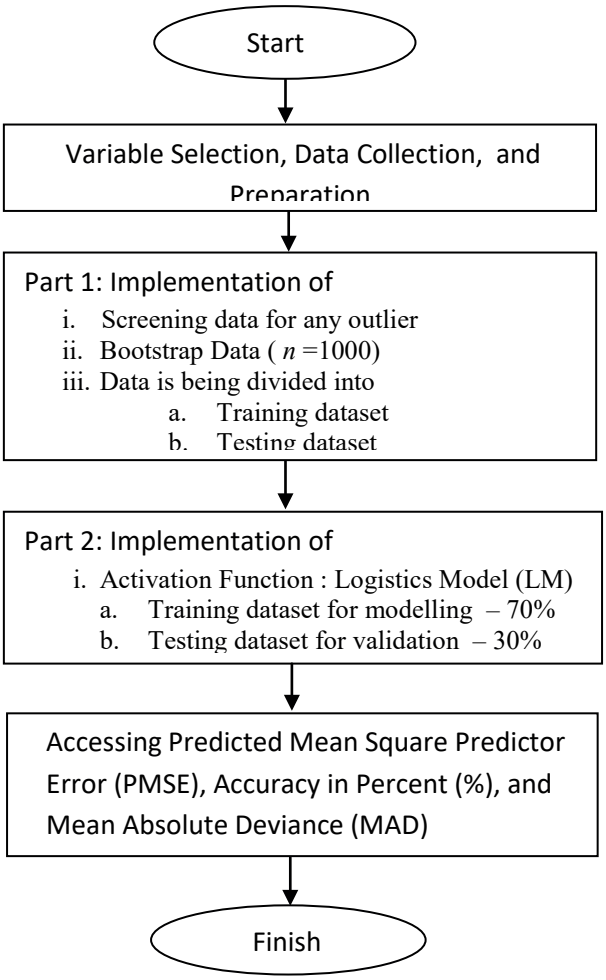

Figure 1. Flowchart of the proposedstatisticallogistic modeling

To illustrate this process, we've included a diagram in Figure 1. Strengths of the study include the fact that it investigates a model that considers clinically relevant variables. Following the preparation of the data, a bootstrapping technique will be developed. The bootstrap method creates a sample of the same size as the original sample, but each observation is repeated several times and others discarded[6,7].

The syntax in R

#/STEP 1-Dataset for Biometry Modeling Study/#

```
Input =("
Marital smoke hyper sys fbs tc hdl alt alp urea
0 2 0 120 5.300 3.40 1.48 10.00 89.00 5.70
0 2 0 110 11.700 6.04 1.29 89.00 67.00 2.90
0 2 \0 130 7.400 1.96 1.51 23.00 40.00 5.70
0 0 0 128 4.600 3.56 1.08 31.00 106.00 3.90
0 2 0 135 6.400 6.04 .97 24.00 51.00 5.20
0 2 0 122 9.800 1.45 .96 22.00 34.00 5.80
0 1 0 134 4.400 3.70 1.15 37.00 96.00 4.50
0 2 0 128 9.900 4.72 .76 30.00 85.00 6.60
0 2 0 130 6.100 4.59 1.09 25.00 57.00 4.70
0 2 0 146 14.700 5.31 1.24 16.00 55.00 13.90
0 1 0 127 4.800 4.35 1.19 16.00 114.00 7.50
0 1 0 130 5.800 3.07 .74 45.00 93.00 27.00
0 2 0 120 6.000 4.64 1.45 42.00 89.00 6.30
0 2 0 120 6.000 4.64 1.45 42.00 89.00 6.30
0 2 0 120 12.200 3.94 .81 89.00 85.00 5.30
0 2 0 116 5.200 3.81 1.07 27.00 85.00 5.20
0 2 0 140 10.200 4.02 .96 27.00 83.00 7.50
0 2 0 130 8.000 4.54 1.17 35.00 78.00 4.30
0 2 1 110 7.600 3.88 1.04 55.00 56.00 3.20
1 2 1 101 8.700 4.09 1.00 41.00 54.00 2.30
1 2 1 101 8.700 4.09 1.00 41.00 54.00 2.30
0 2 0 155 6.800 3.74 1.07 15.00 52.00 6.30
0 2 0 140 7.500 4.93 .93 53.00 96.00 5.20
0 2 1 138 12.100 5.79 .99 58.00 84.00 5.60
0 1 1 114 11.200 4.20 .96 34.00 99.00 4.70
0 2 0 125 7.100 4.95 1.24 35.00 49.00 4.60
0 2 0 138 999.000 4.28 1.09 23.00 135.00 6.60
0 2 1 127 8.100 3.80 .85 45.00 58.00 68.00
0 2 0 120 6.400 5.62 1.22 144.00 77.00 4.20
0 2 0 120 5.400 4.03 .68 20.00 75.00 4.10
0 0 0 140 7.500 5.42 1.03 30.00 63.00 5.30
0 1 0 130 5.800 3.60 1.03 25.00 47.00 4.10
0 1 0 130 5.800 3.60 1.03 25.00 47.00 4.10
0 0 0 130 8.200 3.90 1.22 42.00 99.00 4.90
0 2 0 130 8.000 5.01 1.39 24.00 72.00 5.60
0 0 0 118 5.800 4.71 1.18 15.00 79.00 5.70
0 2 0 130 8.600 5.43 1.29 26.00 172.00 12.10
0 2 0 120 17.300 7.18 1.47 25.00 118.00 4.60
2 2 0 140 3.800 5.13 1.19 9.00 82.00 6.60
")

data1 = read.table(textConnection(Input),header=TRUE)
#/Performing Bootstrap for 1000/#
mydata <- rbind.data.frame(data1, stringsAsFactors = FALSE)
iboot <- sample(1:nrow(mydata),size=1000, replace = TRUE)
data <- mydata[iboot,]

#/Performing Multiple Logistics/#
#/Model Fitting/#
model = glm(hyper ~Marital+smoke+sys+fbs+tc+hdl+alt+alp+urea,
        data=data,family = binomial(link="logit"))
```

```
#/Performing Summary of the Model/#
summary(model)
exp(model$coefficients)
exp(confint(model))

#/Overall p-value For Model/#
anova(model, update(model, ~1), test="Chisq")

#/MultiLayer Perceptron Model/#
#/STEP 2-Install the Neuralnet Package/#
if(!require(neuralnet)){install.packages("neuralnet")}
library("neuralnet")

#/STEP 3- Checking For the Missing Values/#
apply(data, 2, function(x) sum(is.na(x)))

#/STEP 4 - Max-Min Data Normalization/#
normalize <- function(x) {return ((x - min(x)) / (max(x) - min(x)))}
maxmindf <- as.data.frame(lapply(data, normalize))

#/STEP 5-Determine the Training and Testing of the Dataset/#
#/70% for Training and 30% For Testing/#
index = sample(1:nrow(data),round(0.70*nrow(data)))
Training <- as.data.frame(data[index,])
Testing <- as.data.frame(data[-index,])

#/STEP 6-Plotting the Architecture of MLP Neural Network/#
nn <- neuralnet(hyper~Marital+smoke+sys+fbs+tc+hdl+alt+alp+urea,data=Training,
hidden=c(5),act.fct = "logistic",
linear.output = FALSE, stepmax = 1000000)
plot(nn)
options(warn=-1)
nn$result.matrix

#/Testing The Accuracy of The Model- Predicted Result/#
#/STEP 7-Predicted Results Are Compared To The Actual Results/#
Temp_test <- subset(Testing, select = c("Marital","smoke","sys","fbs","tc",
                                        "hdl","alt","alp","urea"))
head(Temp_test)
nn.results <- compute(nn, Temp_test)

#/Results/#
results <- data.frame(actual = Testing$hyper,
prediction = nn.results$net.result)

#/STEP 9-Use The Predicted Mean Squared Error NN (MSE-forecasts the Network)
#/As a Measure of How Far the Predictions Are From The Real Data/#
predicted <- compute(nn,Testing[,1:9])
MSE.net <- sum((Testing$hyper - predicted$net.result)^2)/nrow(Testing)

#/STEP 10-Printing the Predicted Mean Square Error/#
MSE.net

########################Neural Network Parameter Output##########################
#/STEP 11-Neural Network Parameter Output/#
library(neuralnet)
nn <- neuralnet(hyper ~Marital+smoke+sys+fbs+tc+hdl+alt+alp+urea,data=Training,
                hidden=4,act.fct = "logistic", linear.output = FALSE,
 stepmax = 1000000)
nn$result.matrix
############################Model Validation Calculation########################
#/STEP 12- Model Validate/#
results <- data.frame(actual=Testing$hyper,prediction=nn.results$net.result)
results
summary(results)

############################Model Accuracy Calculation ##########################
#/STEP 13- Model Accuracy/#
predicted1=results$prediction*abs(diff(range(data$hyper)))+min(data$hyper)

#/Print(Predicted)/#
actual1=results$actual*abs(diff(range(data$hyper)))+min(data$hyper)

#/Print(Actual1)/#
deviation= ((actual1-predicted1))
#/Print(deviation)/#

#/Mean Absolute Deviance/#
value=abs(mean(deviation))
print(value)
accuracy_in_percent=(1-value)*100
accuracy_in_perc
```

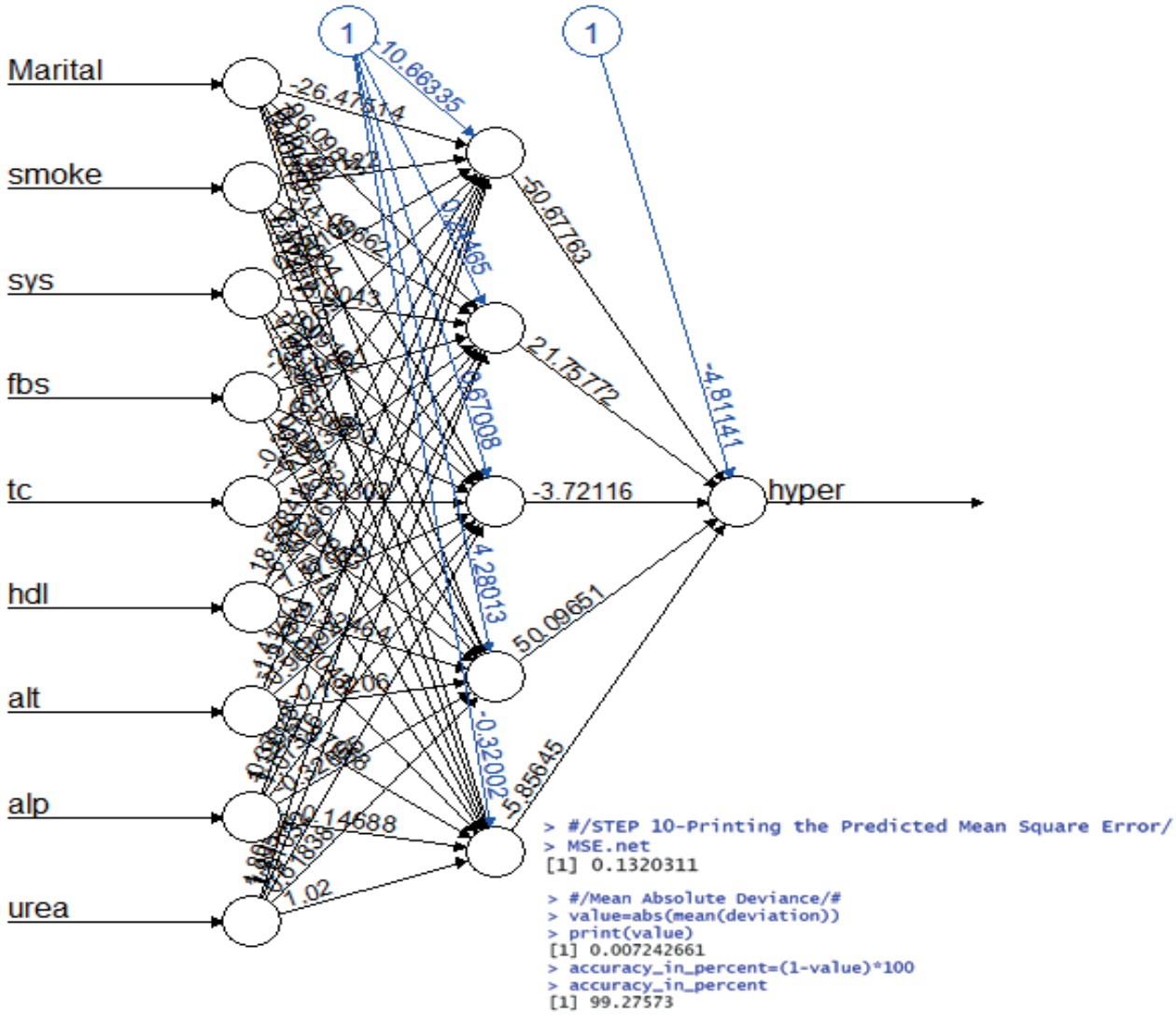

Figure 2. The architecture of the best (M.L.P.) model with nine input variables, one hidden layer, and one output node (Proposed Model)

Results:

This current study aims to investigate the performance of Multilayer perceptron, which is based on activation function: logistics model. This M.L.P.considers both the training and testing datasets. The optimal model for multiple logistic regression was determined by choosing variables that generate the lowest Predicted Mean Square Error as determined by the multiple logistic regression algorithm.

The Result of Regression Modeling

Results of multiple logistic regression using a training dataset are shown in Table 2, where the hypertension status is a dependent variable that serves as the study's outcome. The logistic model's mean absolute deviance (MAD) is 0.00724 in this case, and it indicates how to spread out the data we have. Remarkably similar data will have a small spread. In contrast, wildly dissimilar data will have a large distance between them. In this instance, it speaks to the accuracy and reliability of our forecast data. The train to test split is 70:30, meaning 70% of the data is available for modeling and 30% for testing. Table 2 summarises the results of the multiple regression analysis. Below is the model (Figure 2).

Table 2. Result of Multiple Logistic Regression with combining the bootstrap method training and testing dataset

Variable	Estimate	Std. Error	Z-Value	P-Value
(Intercept)	19.8398118	2.0507974	9.674	< 2e-16*
Marital	1.5390314	0.3292236	4.675	2.94e-06*
Smoke	-0.2298446	0.238863	-0.905	0.33653
Sys	-0.1597623	0.0155866	-10.250	< 2e-16*
Fbs	0.0005294	0.0027233	0.194	0.8459
Tc	0.9926178	0.2018781	4.917	8.79e-07*
Hdl	-5.8695010	0.9052988	-6.483	8.96e-11*
Alt	0.0016224	0.0046291	0.350	0.7260
Alp	-0.0131493	0.0078448	-1.676	0.0937*
Urea	0.0816538	0.0117980	6.921	4.49e-12*

Multiple Logistic Regression was applied
Significant at the level of the 0.25

The variable validation in this section was accomplished by using the established bootstrap method for integrated multiple logistic regression. In this case, nine variables were chosen for analysis: marital status, smoking status, systolic blood pressure, fasting blood glucose, total cholesterol, high-density lipoprotein, alanine transferase, alkaline phosphatase, and urea reading.Six factors were found to have a significant effect on hypertension. Table 3 summarizes the detailed output. Marital status (: 1.183519; p< 0.25), systolic blood pressure (:-0.144516; p<0.25), total cholesterol (:0.9585890; p< 0.25), high-density lipoprotein (:-5.927411; p< 0.25), alkaline phosphatase (:-0.008973; p> 0.25), and urea reading (:0.064169; p< 0.25). Figure 3 show the architecture of the best MLP model with six input variables, one hidden layer, and one output node.

Table 3. Result of Multiple Logistic Regression with combining the bootstrap method training and testing dataset

Variable	Estimate	Std. Error	T-Value	P-Value
(Intercept)	18.071279	1.649753	10.954	< 2e-16*
Marital	1.183519	0.311719	3.797	0.000147*
Sys	-0.144516	0.013025	-11.095	< 2e-16*
Tc	0.9585890	0.156074	6.142	8.15e-10*
Hdl	-5.927411	0.730548	-8.114	4.91e-16*
Alp	-0.008973	0.006054	-1.482	0.138321*
Urea	0.064169	0.011142	5.759	8.45e-09*

Multiple Logistic Regression was applied
Significant at the level of the 0.25

Model evaluation of the model

In this scenario, the model evaluation can be derived from the forecast value. The accuracy of the prediction will be determined by a comparison of the actual and predicted values. The testing dataset will be used to evaluate the model built from the training data set.When comparing actual and predicted data, the distance prediction will be used to determine the difference. The R syntax provides a model assessment approach that can be used to evaluate subsequent methods. Table 4 displays the "Actual" and "Predicted" values obtained from the proposed methodology.

Table 4. The "Actual and "Predicted" value from the proposed methodology

Actual	Predicted
0	0.1639555228
0	0.1639555228
0	0.1639555228
0	0.0009031840
1	0.5306543691

There isn't much of a difference between the "Actual" and "Predicted" values. The results of a paired sample t-test showed no statistically significant differences.

Table 5. Summary of "Actual" and "Predicted" value of the proposed model

Actual	Predicted
Min. :0.00	Min. : 0.0009032
1st Qu.:0.00	1st Qu.: 0.1639285
Median :0.00	Median : 0.1639555
Mean : 0.17	Mean : 0.1456979
3rd Qu.:0.00	3rd Qu. : 0.1639555
Max. :1.00	Max. : 0.5306547

According to the proposed model, the "Actual" and "Predicted" values are shown in Table 5."Actual" and "Predicted" were statistically indistinguishable, according to the findings. This proves the superiority of the proposed model.

Discussion:

We successfully applied the proposed method, which is extremely useful for estimating event probabilities (predict the odds of being a case). The hybrid method, which we obtained and harmonized, produced an extremely accurate and reliable model. Non-linear regression models have several limitations,

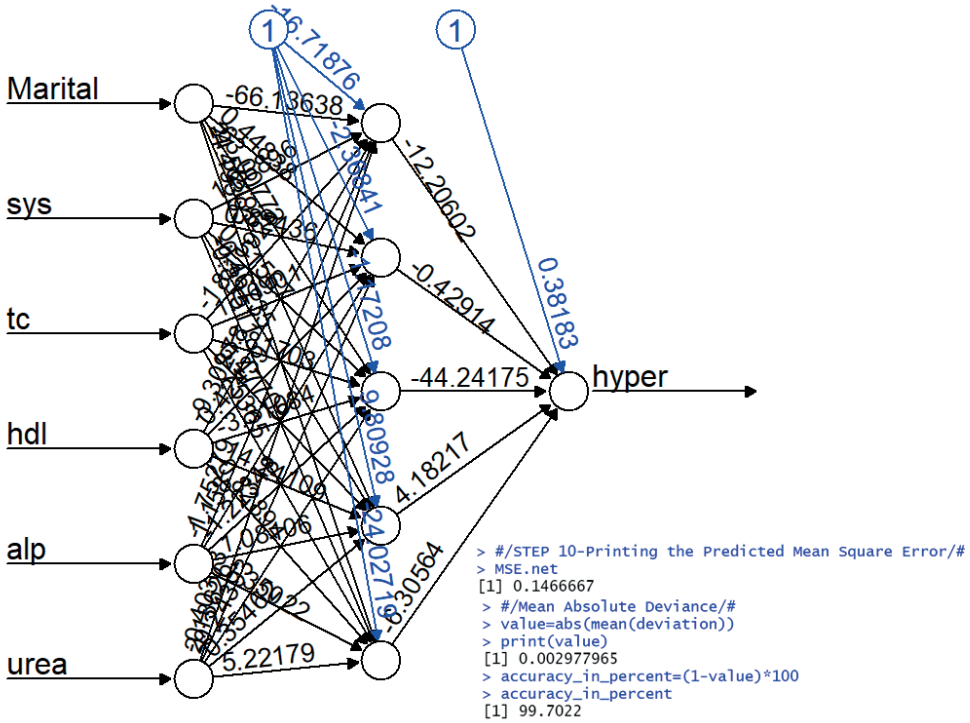

Figure 3. The architecture of the best (M.L.P.) model with six input variables, one hidden layer, and one output node (Obtained model)

including estimation; the calculation procedures are more complicated, less accurate, and less precise for each predictor variable and outcome. The findings revealed that marital status, systolic blood pressure, total cholesterol, high-density lipoprotein, alkaline phosphatase, and ureaare the most critical factors influencing hypertension.

This paper focuses on the development of methodologies for multiple logistic regressions. This paper is primarily concerned with the development of methods for multiple logistic regression. It was the primary goal of the study to develop, test, and validate a regression model. The primary purpose of this project was to combine bootstrapping and MLR techniques for developing and implementing medical statistics strategies. Clinical expert opinion is incorporated into the variable selection process. Using the initial data set, the bootstrap method creates a "mega" file at the beginning of the operation. By contrast, the bootstrap procedure generates a large sample of file replacements. Thirdly, the bootstrap method generates and saves statistical samples. As a final point, the bootstrap method iteratively repeats this process, often thousands of times. Finally, the data is prepared for the following procedure in the fifth stage. The R syntax algorithm enables the application to be integrated with the methodology concept.

The first step is to seek the advice of an expert when selecting variables. Finally, the bootstrap is applied to the dataset. Separate data will be used for training and testing. The R syntax algorithm establishes a link between the application and the notion of method-based methodology.The first step is to choose variables with the support and advice of a medical professional. After that, the data will be subjected to the bootstrap procedure. At this point, 70% of the bootstrap data will be designated as a training dataset, and 30% will be designated as a testing dataset. For both building and testing the model, we will use data from the training dataset. A successful model will have the smallest mean absolute deviance. Syntax was used to calculate the following formula based on actual and predicted values. The study's findings helped the decision-maker get the best possible outcome. Incorporating statistical formulations, computation using R syntax, and the multiple logistic regression package resulted in highly successful

logistic modeling.The most difficult tasks are selecting appropriate input parameters, preparing data for logisticmodeling, and standardizing it.

Conclusion:

New hybrid methods will be developed using bootstrapping and multiple logistic regression in this study. The R syntax for this methodology was designed to ensure that the researcher could fully comprehend the illustration. Hypertension is the dependent variable in this research. At the same time,marital status, smoking status, systolic blood pressure, fasting blood glucose, total cholesterol, high-density lipoprotein, alanine transferase, alkaline phosphatase, and urea reading are the independent variables. Based on the model, factors emerged as the most critical factors. According to regression theory, the more accurate a model is, the higher its R squared value. In addition, the smallest predictive value generated from the obtained model can be used to evaluate the test's fit. As a result, a hybrid model can help us better understand the utility of this method and its relative contribution to the outcome. The statistical strategy proposed in this study in R demonstrates that regression modeling outperforms, and a Mean absolute deviant error is given as 0.00297. According to the study's conclusion, a hybrid model technique proposed in the study is superior to other methods.

Acknowledgment

This study was funded by the Ministry of Higher Education Malaysia (MoHE) for Fundamental Research Grant Scheme (FRGS) with Project Code: FRGS/1/2018/STG06/USM/02/12, and the authors would like to thank Universiti Sains Malaysia (USM).

Conflict of Interest: None

The patient's medical records were reviewed after getting permission from the Director, Hospital USM. Ethical approval was taken from the Human Research Ethics Committee ofUSM (JEPeM) , JEPeM Code: USM/JEPeM/22040245.

Author's contribution:

Data gathering and idea owner of this study: WMAWA, MNA, NAR, FMMG, NAA, ZMYB.

Study design: WMAWA, MNA, NAR, FMMG, NAA, ZMYB.

Data gathering: WMAWA, MNA, NAR, FMMG, NAA, ZMYB.

Writing and submitting manuscript: WMAWA, MNA, NAR, FMMG, NAA, ZMYB and MKA

Editing and approval of final draft: WMAWA, MNA, NAR, FMMG, NAA, ZMYB and MKA

References:

1. Ajila V, Shetty H, Babu S, Shetty V, Hegde S. Human Papilloma Virus Associated Squamous Cell Carcinoma of the Head and Neck. *Journal of Sexually TransmittedDiseases*. 2015; 2015: 1-5. Available from: https://doi.org/10.1155/2015/791024
2. AliJ, SabihaB, JanH, HaiderS, Khan A, Ali S. Genetic etiology of oral cancer. *Oral Oncology*. 2017: **70**(Supplement C): 23-28. Available from: https://doi.org/10.1016/j.oraloncology.2017.05.004
3. Chaturvedi AK. Epidemiology and clinical aspects of HPV in head and neck cancers. Head and Neck Pathology. 2012;**6**(S1):16–24. Available from: https://doi.org/10.1007/s12105-012-0377-0
4. Chaturvedi A, Engels E, Anderson W, Gillison M. Incidence Trends for Human Papillomavirus–Related and –Unrelated Oral Squamous Cell Carcinomas in the United States. Journal of Clinical Oncology. 2008; **26**(4): 612-619. Available from: https//doi.org/10.1200/jco.2007.14.1713
5. Cheong S, Vatanasapt P, Yi-Hsin Y, Zain R, Kerr A, Johnson N. Oral cancer in South East Asia. Translational Research in Oral Oncology. 2017; **2**: 2057178X1770292.
6. Efron B. The jackknife, the bootstrap and the other resampling plans. Philadelphia, Pa., Society for industrial and Applied Mahtematics; 1982.
7. Efron B, Tibshirani R. An introduction to the bootstrap. Boca Raton, Fla.: Chapman & Hall/CRC; 1998.
8. Fan S, Tang Q, Lin Y, Chen W, Li J, Huang Z et al. A review of clinical and histological parameters associated with contralateral neck metastases in oral squamous cell carcinoma. *International Journal of Oral Science*. 2011; **3**(4): 180-191. Available from: https//doi.org/10.4248/ijos11068
9. Frithiof L, Wersäll J. Virus-like particles in papillomas of the human oral cavity. *Archives of Virology*. 1967;**21**(1):31-44.
10. Goot-Heah K, Kwai-Lin T, Froemming G, Abraham M, Rosdy N, Zain R. Human Papilloma Virus 18 Detection in Oral Squamous Cell Carcinoma and Potentially Malignant Lesions Using Saliva Samples. *Asian Pacific Journal of Cancer Prevention.* 2012; **13**(12): 6109-6113.
11. SungH, FerlayJ, SiegelRL, LaversanneM, SoerjomataramI,malA, BrayF.Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*. 2021; **71**(3): 209-249.
12. JiangS, DongY.Human papillomavirus and oral squamous cell carcinoma: A review of HPV-positive oral squamous cell carcinoma and possible strategies for future. Curr Probl Cancer. 2017; **41**(5): 323-327. Available from: https//doi.org/10.1016/j.currproblcancer.2017.02.006
13. JensonA, LancasterW, HartmannD, Shaffer JrE. Frequency and distribution of papillomavirus structural antigens in verrucae, multiple papillomas, and condylomata of the oral cavity. *The American Journal of Pathology*. 1982; **107**(2): 212.
14. KearnetPM, WheltonM, ReynoldsK, MuntnerP, WheltonPK, HeJ.Global burden of hypertension: analysis of worldwide data. *Lancet*. 2005; **365**(9455): 217-223.
15. KerishnanJP, GopinathSC, KaiSB, TangT-H, Ng HL-C, RahmanZAA, HashimU, ChenY.Detection of human papillomavirus 16-Specific IgG and IgM antibodies in patient sera: A potential indicator of oral squamous cell carcinoma risk factor. *International Journal of Medical Sciences*. 2016; **13**(6): 424.
16. KrügerM, PabstA, WalterC, SaghebK, GüntherC, BlattS, WeiseK, Al-NawasB, ZiebartT. The prevalence of human papilloma virus (HPV) infections in oral squamous cell carcinomas: a retrospective analysis of 88 patients and literature overview. *Journal of Cranio-Maxillofacial Surgery*. 2014; **42**(7): 1506-1514.
17. MacMahonS, AldermanMH, LindholmLH, LiuL, SanchezRA, SeedatYK.Blood-pressure-related disease is a global health priority. *Lancet*.2008; **371**(9623): 1480-1482.
18. Mohamad Ghazali FM, W Ahmad WM, Srivastava KC, Shrivastava D, Mohd Noor NF, Nizam Akbar NA, Aleng NA, Alam MK.A study of creatinine level among patients with dyslipidemia and type 2 diabetes mellitus using multilayer perceptron and multiple linear regression. *J Pharm Bioallied Sci*. 2021; **13**(Suppl S1):795-800.
19. PoorolajalJ,FarbakhshF,MahjubH,BidarafshA,BabaeeE. How much excess body weight, blood sugar, or age can double the risk of hypertension?. *Public Health.* 2016; **133**: 14-18.
20. TaskinenMR. Diabetic dyslipidemia. Atherosclerosis Supplements. 2002; **3**(1): 47-51.
21. Vargas-FerreiraF, NedelF, EtgesA, GomesAPN, FuruseC, TarquinioSBC.Etiologic factors associated with oral squamous cell carcinoma in non-smokers and non-alcoholic drinkers: a brief approach. *Brazilian Dental Journal*. 2012; **23**: 586-590.
22. VinodMahatoR, GyawaliP, RautPP, RegmiP, SinghKP, PandeyaDR,GyawaliP.Association between glycaemic control and serum lipid profile in type 2 diabetic patients: Glycated haemoglobin as a dual biomarker. *Biomedical Research*. 2011; (0970-938X): **22**(3).