A VARIABLE SELECTION IN ORDERED LOGISTIC REGRESSION MODEL USING DECISION TREE ANALYSIS FOR THE CLASSIFICATION: A CASE STUDY OF HYPERTENSION MODELING

Wan Muhamad Amir W Ahmad[1*], Hazik Bin Shahzad[1], Mohamad Nasarudin Adnan[1], Farah Muna Mohamad Ghazali[1], Noraini Mohamad[1] Norhayati Yusop[1]**, Nor Farid Mohd Noor[2], Nor Azlida Aleng[3], Mohamad Shafiq Mohd Ibrahim[4]**

[1*]School of Dental Sciences, Health Campus. Universiti Sains Malaysia (USM), 16150 Kubang Kerian, Kota Bharu, Kelantan, Malaysia.
[2] Faculty of Medicine, Universiti Sultan Zainal Abidin (UniSZA), Medical Campus, Jalan Sultan Mahmud, 20400 Kuala Terengganu, Terengganu, Malaysia.
[3] Faculty of Ocean Engineering Technology and Informatics, Universiti Malaysia Terengganu (UMT), 21030 Kuala Nerus, Terengganu, Malaysia.
[4]Kulliyyah of Dentistry, International Islamic University Malaysia, IIUM Kuantan Campus, Jalan Sultan Ahmad Shah, Bandar Indera Mahkota, 25200 Kuantan, Pahang, Malaysia

ABSTRACT

Background and Objective: Hypertension is a public health issue that depicts high blood pressure in which the force of the blood vessels increases persistently. According to the WHO, one in four men and one in five women have hypertension. Twenty to thirty percent of the adult population and more than five to eight percent of pregnancies worldwide suffer from hypertension, which is frequently curable when detected and treated early. Recognizing the significance of statistical modeling in hypertension, this study aims to develop a method that stakeholders can use to predict and manage hypertension cases. Two approaches used in this study were decision tree and ordinal regression. Both methods will be harmonized in the R syntax with some modification and extension. **Materials and Methods:** In this paper, we developed the method of decision tree analysis using R syntax with embedding the prediction classification. The classification for prediction with accuracy will indicate the successful classification analysis. This study used hypertension data consisting of one thousand observations to illustrate the development method. Before further testing, each pre-selected variable's clinical relevance and significance will be evaluated. Four selected variables will be tested using the decision tree. The selected variables are blood pressure, glucose, height, and triglycerides. The classification obtained will be used as input for the ordinal regression modeling. **Result:** It has been found that the level of hypertension can be determined by systolic blood pressure, glucose, and triglycerides, according to the most recent published research. These four variables are chosen and used for the input of the ordinal regression. The suggested variables will apply to the ordered logistic regression, and the goodness of measurement is conducted using the developed syntax. The significance level is set at 0.05 level. **Conclusion:** We can conclude that our proposed method yields excellent results with the highest level of forecasting precision possible. The method approach provides an accurate evaluation of the fit of the final model. The superior performance of the model led to improved outcomes and effective decision-making management.

Keywords: Hypertension, decision tree analysis, ordinal logistic regression

1.0 INTRODUCTION

Hypertension, often known as high blood pressure, is a common health problem that significantly impacts the world's health. According to studies on the clustering of cardiovascular disease burden in 21 regions, hypertension-related consequences resulted in almost 9 million fatalities. Approximately forty percent of individuals older than 25 have been diagnosed with hypertension [6,7]. The excellent study of the N.C.D. Risk Factor Collaboration on changes in blood pressure around the world from 1975 to 2015, which comprised 1,479 population-based research, found that the number of adults with hypertension increased from 594 million in 1975 to 1.13 billion in 2015 [1].

Furthermore, according to the WHO, nearly one billion adults worldwide had hypertension in 2000, and nearly 1.56 billion people (29.2% of the global population) are projected to have hypertension by 2025 [9, 12]. Uncontrolled and untreated hypertension is associated with an increased risk of total and cardiovascular death in the general hypertensive population [10]. Finding an effective strategy to control high blood pressure has been a global problem, even though it has been reported as an important public health issue [15].

The treatment is also less effective because it is administered only at the end stages of hypertension [16]. It has been found that people with myocardial infarction are 2.5 times more likely than those with normal blood pressure, regardless of ethnicity, sex, or smoking status, to have extremely high blood pressure [13]. Prehypertensive patients were 1.5 times more likely to develop cardiovascular disease than those with normal blood pressure. [5, 11]. More than 180 million Chinese people had hypertension in 2000, which is expected to rise by another 100 million in 2025. Asia was home to over 40% of the 1.13 billion adults with high blood pressure in 2015 [2, 4] with 226 million living in China alone. People in the highest blood pressure quintile of the Framingham Heart Study cohort were more than twice as likely to develop congestive heart failure as those in the lowest quintile [8]. According to population-based research, men, older people, and those with low household incomes are more likely to suffer from hypertension in Malaysia. Obesity, smoking, excessive alcohol consumption, high cholesterol, and diabetes are other predisposing factors [3]. Studies on clinical hypertension have focused on adults already suffering from the disease. Still, there has been a lack of attention paid to factors contributing to the severity of hypertension, particularly among the hypertensive population.

Consequently, the current study seeks to model the ordinal regression using the result obtained from decision tree analysis. The ultimate objective of this research is to develop the R syntax for ordinal logistic regression modeling purposes considering decision tree analysis. In the future, it is anticipated that this programming will facilitate optimal decision-making outcomes for the researcher.

2.0 MATERIAL AND METHODS

Decision Tree

Decision Trees are a kind of Supervised Machine Learning in which data is continually partitioned based on a given parameter. Two entities may be used to describe the tree: decision nodes and leaves. The leaves represent decisions and results, and the data is separated at the decision nodes. A decision tree is an excellent and efficient method for classification, prediction, and facilitating decision-making in sequential decision problems [17]. This strategy has been widely employed in numerous fields. For instance, in the medical field, the decision-maker frequently encounters a sequential decision problem involving choices that, depending on chance, result in various outcomes. Figure 1 shows that decision trees are the best way to graphically show this kind of information. It has helped stakeholders to understand problems intuitively and make better decisions while also assisting in how decisions can be made and what could happen. A decision tree (Fig.1) consists of three types of nodes (a) decision node (b) chance node (c) Endpoint node/Terminal node [14].

In this study, a data item x is a vector of d attribute values with an optional class label y. We denote the set of attributes as A (set of attributes) = $\{A_1, A_2, ..., A_d\}$. Thus, we can characterize x as $\{x_1, x_2, ..., x_d\}$, where $x_1 \in A1$, $x_2 \in A2$, ..., $x_d \in A_d$. Let Y (domain of class values) = $\{y_1, y_2, ..., y_m\}$ be the set of class labels. Each training item x is mapped to a class value y where $y \in Y$. The complete set of training data is X. (set of training data). A partitioning rule S (splitting rule) subdivides data set X into a set of subsets collectively known as X_S; that is, $X_S = \{X_1, X_2,...,X_k\}$ where $\cup i\ X_i = \mathbf{X}$. A decision tree is a rooted tree in which each set of parent nodes corresponds to a partitioning (X_S) of the parent's data set, with the full data set associated with the root. The number of items in X_i that belong to class y_j is $|\boldsymbol{X}_{ij}|$. The probability that a randomly selected member of X_i is of class y_j is $p_{ij} = |X_{ij}| / |X_i|$

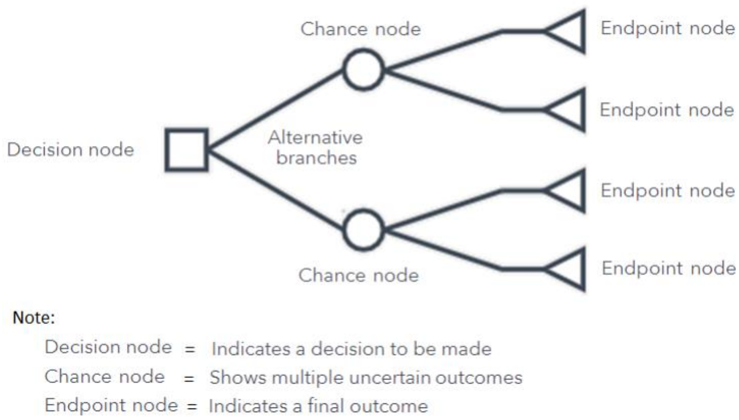

Figure 1: Decision trees are graphical models for describing Sequential decision problems

Conditional Inference Trees (Ctree)

The developed syntax uses the conditional inference trees (Ctree) in the R package party kit. A conditional inference tree CTree was proposed by [14]. CTree divides the process of splitting into two discrete parts. First, the variable to split is identified based on a measure of the relationship between each covariate and the result of interest. After determining the splitting variable, the optimal split point for that variable is calculated. At each step in the splitting process, CTree uses formal statistical inference procedures. In a regression model, the coefficient is used to measure the relationship between each covariate and the outcome. A node is only split if there is enough evidence to reject the global null hypothesis, which says that none of the covariates has a one-to-one relationship with the outcome. If the global null hypothesis is rejected, the covariate with the strongest link to the outcome of interest is chosen as a candidate for splitting. No variable is chosen for splitting if the minimal p-value exceeds the multiplicity-adjusted significance threshold, and the node is labeled as a terminal node. The "conditional" refers to the fact that after the initial split, subsequent inference takes place inside subgroups, i.e., conditional on subgroup membership; Ctree-based splitting choices on marginal regression models. Four selected variables are glucose levels, patient height, triglyceride level, and systolic blood pressure [14].

ISSN 2515-8260 Volume 10, Issue 01, 2023

Bootstrap

Bootstrap begins by selecting a sample at random from the population and then computing sample statistics. After multiple iterations of the initial samples, the bootstrap generates a pseudo population by substituting samples. Random sampling with substitution produces samples that are not identical to the original sample. The bootstrap computes statistics for each sample as the sample is drawn with replacement.

Ordinal Regression

When modeling a categorical dependent variable (one with more than two categories) as a function of one or more independent variables, ordinal regression can be helpful. Ordinal logistic regression (OLR) is a type of logistic regression analysis used when the response variable contains more than two categories with a natural rank or order. In ordinal logistic regression, we must have an ordinal-scaled dependent variable. The hypertension reading has a ratio measurement scale; therefore, we must convert it to a three-scaled ordinal variable. After converting the hypertension reading to an ordinal scale, we will fit the ordinal regression. The maximum likelihood method will be used to estimate the regression parameter's value.

The model for ordinal is given by

$$y_i^* = x_i\beta + \varepsilon_i \tag{1}$$

The dependent variable, however, is categorized, so we must use:

$$C_x(x) = \ln\left[\frac{P(Y \le j \mid x)}{P(Y > j \mid x)}\right]$$

and $\ln\left(\frac{\sum pr(event)}{1-\sum pr(event)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 +, \ldots, + \beta_k X_k$. It can be

summarized as

$$\ln\left(\frac{P(Y \le j \mid x)}{1 - P(Y \le j \mid x)}\right) = \alpha_j + \beta_i X_k \qquad i = 1...k, \quad j = 1, 2, ..., p-1 \tag{2}$$

where

α_j = called threshold or intercept

β_i = Parameter in the model

X_{i1} = Set of factors or independent variables.

Equation (2) above is an ordinal logistic model for *k* predictors with the *p*-1 levels response variable [18].

DATA AND R SYNTAX

The research was conducted at Hospital Universiti Sains Malaysia (USM) in Kubang Kerian, Kelantan, Malaysia. A total of 1000 patients took part in this study. The data summary for the selected variable in the analysis is described in Table 1.

ISSN 2515-8260 Volume 10, Issue 01, 2023

Table 1: Data description of research variables

Code-variables	Explanation of user Variables
Glucose	Glucose Level
Height	Patient Height
Trig	Triglycerides level
Systolic	Systolic Blood Pressure Reading
Kclass	Hypertension level
	0 = Normal, 1= Borderline and 2 = Hypertension

The statistical analysis was conducted using the developed R Studio software package methodology. This study combines three statistical techniques in a single syntax for optimal research results.

Bootstrapping Ordinal Regression Modeling with Decision Tree Analysis Using R Syntax

```
#install.packages("party")
library(party)
#install.packages("partykit")
library(partykit)
#install.packages("caret")
library(caret)
#install.packages("class")
library(class)
#install.packages("tree")
library(tree)
#install.packages("caTools")
library(caTools)

Input =("
glucose height trig sysbp hyper kclass
114 178 168 137 definite 2
109 181 332 155 borderline 1
153 183 304 130 definite 2
92 178 81 121 normotensive 0
94 175 98 114 normotensive 0
109 183 471 132 definite 2
107 182 145 160 definite 2
103 173 96 135 normotensive 0
109 170 261 116 normotensive 0
117 167 217 149 definite 2
⋮      ⋮      ⋮
96 172 73 163 definite 2
142 177 70 136 normotensive 0
91 161 82 154 borderline 1
99 167 102 116 normotensive 0
")
data = read.table(textConnection(Input),header=TRUE)
head(data)
```

ISSN 2515-8260 Volume 10, Issue 01, 2023

#To Convert Factors To Numbers
```
data$kclass<-as.factor(data$kclass)
```

#STEP 1: Decision Tree Using The Whole Data
```
dtm <- ctree(kclass~glucose + height + trig + sysbp, data=data)
plot(dtm)
```

#Calculating The Prediction For The Test
```
pred = predict(dtm, data[,-5])
confusionMatrix(pred,as.factor(data$kclass))

##########################################################################
```

STEP 2: Modeling Ordinal Model
```
#if(!require(MASS)){install.packages("MASS")}
library(MASS)
#if(!require(ordinal)){install.packages("ordinal")}
library(ordinal)
#if(!require(erer)){install.packages("erer")}
library(erer)
```

#Performing Bootstrap for 10000
```
mydata <- rbind.data.frame(data, stringsAsFactors = FALSE)
iboot <- sample(1:nrow(mydata),size=10000, replace = TRUE)
Bootdata <- mydata[iboot,]
```
#Converts a Column From Numeric To Factor.
```
data$kclass<-factor(data$kclass)
model  <- clm(kclass~glucose + height + sysbp, data = Bootdata)
options(warn=-1)
summary(model)
options(warn=-1)
anova(model, type="II")
```

Fit Ordered Logit Model And Store Results 'm'
```
m <- polr(kclass~glucose + height + sysbp, data = Bootdata, Hess = TRUE)
```

View A Summary of The Model
```
summary(m)
```

Store Table
```
(ctable <- coef(summary(m)))
```

Calculate And Store The p Values
```
p <- pnorm(abs(ctable[, "t value"]), lower.tail = FALSE) * 2
```

Combined The Table
```
(ctable <- cbind(ctable, `p value` = p))
```

ISSN 2515-8260 Volume 10, Issue 01, 2023

(ci <- confint(m)) # Default Method Gives Profiled of the 95% Confidence Intervals

Calculation Of The Odds Ratios
exp(coef(m))

Combined The Table of Odd Ratio and Confidence Intervals
exp(cbind(OR = coef(m), ci))
#Finished

In this case, four selected independent variables are glucose, height, triglycerides, and systolic blood pressure. The decision tree analysis was used to select the variables in the ordinal regression modeling. It was discovered that glucose, height, and systolic blood pressure all play a role in hypertension classification.

RESULTS

Decision Tree Result

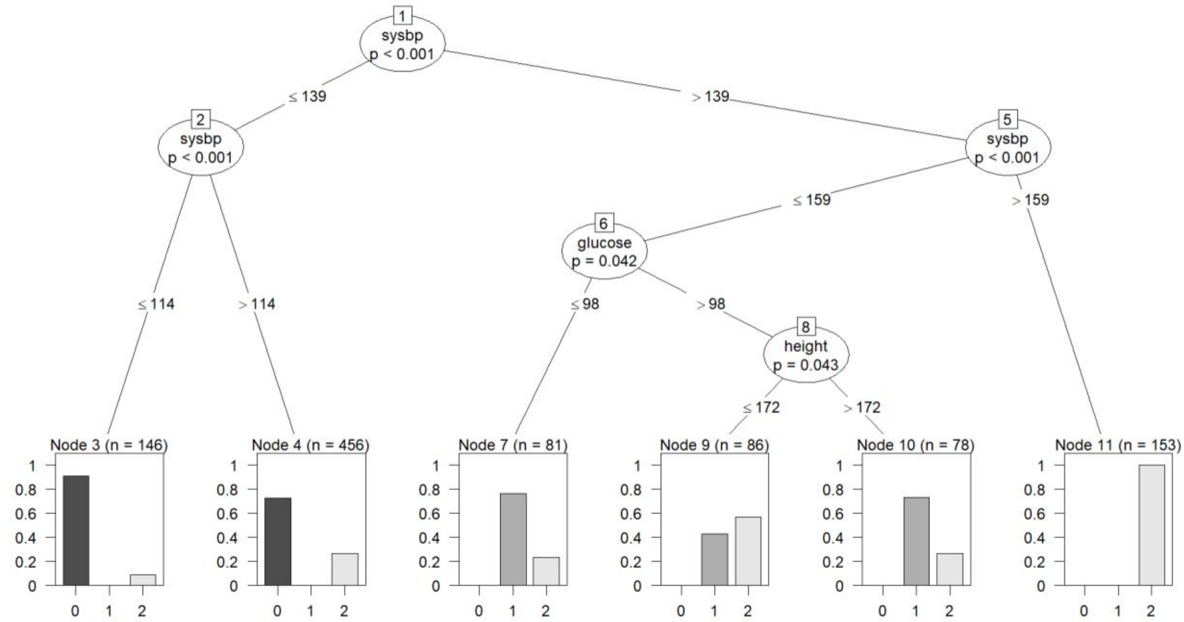

Figure 2: The Decision Tree Analysis

Figure 2 shows the result of the developed decision tree analysis, and three variables were involved in this classification: systolic blood pressure ($p < 0.05$), glucose ($p < 0.05$), and height ($p < 0.05$). These variables will be a feeder for the ordinal regression modeling.

Prediction Using the Result of Decision Tree Result

Table 2: Confusion matrix and Statistics

Reference

	Normal	Borderline	Hypertension
Normal	465	2	135
Borderline	0	119	40
Hypertension	0	37	202

Table 3: Overall Statistics for Decision Tree Analysis

Statistics	Value
Accuracy	: 0.786
95% CI	: (0.7593, 0.811)
P-Value	: < 2.2e-16
Kappa	: 0.6462
McNemar's Test P-Value	: < 2.2e-16

Table 4: The Summary of Statistics by Class

	Class :0	Class :1	Class:2
Sensitivity	1.0000	0.7532	0.5358
Specificity	0.7439	0.9525	0.9406
Pos Pred Value	0.7724	0.7484	0.8452
Neg Pred Value	1.0000	0.9536	0.7700
Prevalence	0.4650	0.1580	0.3770
Detection Rate	0.4650	0.1190	0.2020
Detection Prevalence	0.6020	0.1590	0.2390
Balanced Accuracy	0.8720	0.8528	0.7382

Table 5: Parameter estimate on the ordinal logistic regression model.

Response	**Odds Ratio**	**Coefficient**	**St. Error**	**t-value**	**p-value**	**95% Confidence Interval** Lower	Upper
Cut1		7.4705	0.6713	11.1287	9.096e-29*	-	-
Cut2		8.4823	0.6726	12.6106	1.844e-36*	-	-
Glucose	1.0089	0.0089	0.0008	10.5458	5.309e-26*	0.00727	0.01059
Height	0.9724	-0.0279	0.0036	-7.6897	1.474e-26*	-0.03502	-0.02082
Sysbp	1.0878	0.08424	0.0015	54.576	0.000e-00*	0.08124	0.08729

Multiple logistic regression was applied
**Significant at the level of 0.05*

Table 5 summarizes the results of an ordinal regression model incorporating ordinal regression and decision tree analysis inputs. Table 5 displays the regression with various intercepts and a single cut-off. The intercepts and the cut point can be used to calculate the predicted probabilities of a patient with a given set of characteristics belonging to a specific category. The proposed ordered logistic regression models for the different cut-off points shall be different and represented by a separate equation, so the formulations for the first and second category becomes (the estimated model):

Ordinal logistic regression for Cut 1

$$\text{Logit}\,(P(Y \leq 1)) = 7.4704 + (0.008 \times \text{ Glucose level}) + (-0.026 \times \text{Height}) + (0.084 \times \text{ Systolic Blood Pressure}) \quad (1)$$

Ordinal logistic regression for Cut 2

$$\text{Logit}\,(P(Y \leq 2)) = 8.4823 + (0.008 \times \text{Glucose level}) + (-0.026 \times \text{Height}) + (0.084 \times \text{ Systolic Blood Pressure}) \quad (2)$$

Table 5 displays the results of ordered logistic regression for the significant variables selected with decision tree modeling. The ordinal model considers three factors that adjust the class odds ratio, including glucose levels, height, and systolic blood pressure. Both glucose and systolic blood pressure show that when either increases, there is a higher risk of hypertension. Height has the opposite effect, and with increased height, there are significantly fewer odds of being in a higher category of hypertension.

DISCUSSION

The discussion of this suggested model may be divided into two phases: first, the discussion focuses on the development of the technique, and second, the discussion focuses on the findings, which contribute to the study's findings. The hybrid model, comprised of an ordinal regression model, is constructed based on the model assumption of parallel lines for all corresponding outcome categories. The suggested methodology for an ordered logistic regression model with decision tree analysis requires the premise that the response variable is ordinal and the bootstrap technique's presence increases accuracy. This proposed method can serve as an alternate way to ordinal regression modeling. This proposed method can serve as an alternate way to ordinal regression modeling.

As expenditures for healthcare on non-communicable chronic diseases increases, the significance of predictive modeling for predicting the occurrence of hypertension is essential for the public health field. This study makes it easier for health workers to identify individuals at high risk for hypertension. The proposed method is beneficial for the highest level of forecasting precision possible. The purpose of this paper was to discuss the development of methodologies for ordinal regression. The study’s primary goal was to create, test, and validate a combination of the decision tree, bootstrap, and ordinal regression for developing and implementing medical statistic strategies. Initially, the bootstrap method generates an extremely large dataset. In this instance of research, the decision tree method accurately evaluates variables to be selected with care for the final model. With the increase in routine monitoring, the volume of data has increased dramatically, and numerous variables might be of marginal importance. Alternatively, decision trees can be used to select the most pertinent input variables that will be used to build the models, which can then be used to formulate a clinical hypothesis and inform future researchers. The variable selection derived from the decision tree demonstrates the viability of developing such prediction models; this strategy is highly beneficial for health services planning. This strategy leads to an accuracy rate of 87%, with above 75% positive and negative prediction values. The most significant variables were glucose levels, the height of individuals, and systolic blood pressure. This process can be used as an alternative to ordered regression modeling in cases where the selection of appropriate variables was based on computational analysis that forecasted the importance of the independent variable that should be chosen for the final model. The most perplexing task of any research is selecting the appropriate input parameters, made easy by this methodology. To determine the efficacy of the developed method, the predictive model is applied to actual data, and the resulting output is compared to the actual data. The findings assisted the decision-maker in achieving the best possible outcomes.

Multiple studies in the last decade have investigated the risk factors for hypertension using decision trees. Integrating a decision tree with ordered regression analysis gives us the benefit of creating a more robust clinical application of risk factors. Chang et al. [19] tried a different mining tool for predicting risk factors of hypertension and showed significant results for triglycerides, creatinine, age, and uric acid. Akdağ et al. [20] used decision trees and found the risk factors of hypertension as BMI, waist-to-hip ratio, gender, and triglycerides. In a study in Qatar by AlKaabi et al. [21], comparable results through the random forest and logistic regression analysis linked to age and physical activity, consumption of fruits and vegetables, and history of diabetes as essential predictors of hypertension. Zekewos et al. [22] in Ethiopia also linked systolic blood pressure, age, and BMI to risk factors for developing hypertension. In a longitudinal study, Dimitriadis et al. [23] showed a significant association of hypertension with risk factors of age, gender, and blood glucose levels. Using decision tree analysis, more conclusive, detailed, and reliable results can be obtained. The proposed strategy and acquired results demonstrate the superiority of the modeling technique used. This article aims to develop a hybrid method incorporating bootstrapping, decision trees, and ordered logistic regression. The R syntax provided was designed to ensure that other researchers could follow the steps in the future. In this study, glucose levels, height, and systolic blood pressure were the independent variables. Policymakers and health professionals can utilize this methodology in establishing new programs by updating current preventive strategies. Using bootstrap, decision tree, and ordered logistic regression, predictive models can generate robust diagnostic parameters by producing accurate predictions using available data. When using a decision tree, ordered regression predicts the risk of hypertension by learning from the data and validating the prediction. These models can be constantly customized as prevention strategies to develop the said preventive strategies.

CONCLUSION

The R syntax synchronizes and harmonizes the concept of a methodology-based approach with the application in the first section of the developed methodology. The second part of the methodology is modeling an ordered logistic regression. This research output shows the result and its application to risk factors associated with hypertension. The decision tree ranks the importance of key risk factors. The sensitivity and specificity values for the model were above 75%. We can conclude that our proposed method yields excellent results with the highest forecasting precision possible.

The method approach provides an accurate evaluation of the fit of the final model. The superior performance of the model led to improved outcomes and effective decision-making management. Hopefully, this information will assist a clinician in managing and educating patients about the risk factors associated with hypertension. The R syntax algorithm integrates the idea of a methodology-based approach to finding risk factors, an ongoing process that greatly aids in managing hypertensive patients. The parameters of glucose levels, height, and systolic blood pressure significantly contributed to a patient's overall hypertensive status, proving that the methodology and findings were successfully developed.

Acknowledgments: The authors would like to thank the Ministry of Higher Education Malaysia for Fundamental Research Grant Scheme with Project Code: FRGS/1/2022/STG06/USM/02/10 and Universiti Sains Malaysia (USM).

REFERENCES

1. A Rashid K, KA Narayan, & Azizah Hj Ab Manan. (2008). The Prevalence of Hypertension among the Elderly in Fourteen Villages in Kedah, Malaysia. *Malaysian Journal of Medicine and Health Sciences, 4(2)*, 33-39.
2. Jo, I.G., Ahn, Y., Lee, J., Shin, K.R., Lee, H., & Shin, C. (2001). Prevalence, awareness, treatment, control and risk factors of hypertension in Korea: the Ansan study. *Journal of Hypertension, 19*, 1523-1532.
3. Kearney, P.M., Whelton, M., Reynolds, K., Muntner, P.M., Whelton, P.K., & He, J. (2005). Global burden of hypertension: analysis of worldwide data. *The Lancet, 365*, 217-223.
4. Lawes, C.M., Rodgers, A., Bennett, D.A., Parag, V., Suh, I., Ueshima, H., & MacMahon, S. (2003). Blood pressure and cardiovascular disease in the Asia Pacific region. *Journal of Hypertension, 21(4)*, 707-716.
5. MacMahon, S., Alderman, M.H., Lindholm, L.H., Liu, L., Sánchez, R.A., & Seedat, Y.K. (2008). Blood-pressure-related disease is a global health priority. *The Lancet, 371*, 1480-1482.
6. National Nutrition Monitoring Bureau, National Institute of Nutrition. (2006). Diet & nutritional statu of population and prevalence of hypertension among adults in rural areas. NNM Technical Report No. 24. Hyderabad: National Institute of Nutrition.
7. Kearney, P.M., Whelton, M., Reynolds, K., Whelton, P.K., & He, J. (2004). Worldwide prevalence of hypertension: a systematic review. *Journal of Hypertension, 22*, 11-19.
8. Poorolajal, J., Farbakhsh, F., Mahjub, H., Bidarafsh, A., & Babaee, E. (2016). How much excess body weight, blood sugar, or age can double the risk of hypertension? *Public health, 133*, 14-8.
9. Primatesta, P., & Poulter, N.R. (2006). Improvement in hypertension management in England: results from the Health Survey for England 2003. *Journal of Hypertension, 24(6),* 1187–1192.
10. Singh, R., Suh, I., Singh, V., Chaithiraphan, S., Laothavorn, P., Sy, R., Babilonia, N., Rahman, A., Sheikh, S., Tomlinson, B., & Sarrafzadigan, N. (2000). Hypertension and stroke in Asia: prevalence, control and strategies in developing countries for prevention. *Journal of Human Hypertension, 14*, 749-763.
11. Wang, T.J., & Vasan, R.S. (2005). Epidemiology of uncontrolled hypertension in the United States. *Circulation, 112(11),* 1651-62.
12. WHO/ISH Writing Group. (2003). World Health Organization/International Society of Hypertension (ISH) Statement on Management of Hypertension.
13. Wong, N.D., Black, H.R., Gardin, J.M. (2005). Preventive Cardiology: A Practical Approach Chicago, Mc Graw Hill.
14. Hothorn, T., & Zeileis, A. (2015). Partykit: a modular toolkit for recursive partytioning in R. *J. Mach. Learn. Res., 16*, 3905-3909.
15. Carey, R. M., Muntner, P., Bosworth, H. B., & Whelton, P. K. (2018). Prevention and Control of Hypertension: JACC Health Promotion Series. *Journal of the American College of Cardiology, 72(11)*, 1278–1293. https://doi.org/10.1016/j.jacc.2018.07.008
16. Schrader, J., Lüders, S., Kulschewski, A., Hammersen, F., Plate, K., Berger, J., Zidek, W., Dominiak, P., & Diener, H.C. (2005). Morbidity and Mortality After Stroke, Eprosartan Compared With Nitrendipine for Secondary Prevention: Principal Results of a Prospective Randomized Controlled Study (MOSES). *Stroke, 36(6)*, 1218-1226.
17. Mesaric, J., & Šebalj, D. (2016). Decision trees for predicting the academic success of students. *Croatian Operational Research Review, 7*, 367-388.
18. Adeleke, K.A., & Adepoju, A.A. (2010). Ordinal Logistic Regression Model: An Application to Pregnancy Outcomes. *Journal of Mathematics and Statistics, 6(3)*, 279-285.

ISSN 2515-8260 Volume 10, Issue 01, 2023

19. Chang, C., Wang, C., & Jiang, B.C. (2011). Using data mining techniques for multi-diseases prediction modeling of hypertension and hyperlipidemia by common risk factors. *Expert Syst. Appl., 38(5)*, 5507-5513.
20. Akdağ, B., Fenkçi, S.M., Degirmencioglu, S., Rota, S., Sermez, Y., & Çamdeviren, H. (2006). Determination of risk factors for hypertension through the classification tree method. *Advances in Therapy, 23(6)*, 885-892.
21. AlKaabi, L. A., Ahmed, L. S., Al Attiyah, M. F., & Abdel-Rahman, M. E. (2020). Predicting hypertension using machine learning: Findings from Qatar Biobank Study. *PloS one, 15(10)*, e0240370. https://doi.org/10.1371/journal.pone.0240370
22. Zekewos, A., Egeno, T., & Loha, E. (2019). The magnitude of hypertension and its risk factors in southern Ethiopia: A community based study. *PloS one, 14(8)*, e0221726. https://doi.org/10.1371/journal.pone.0221726
23. Dimitriadis, G.D., Mitrou, P., Lambadiari, V.A., Maratou, E., & Raptis, S.A. (2011). Insulin effects in muscle and adipose tissue. *Diabetes research and clinical practice, 93 Suppl 1*, S52-9.