

ORIGINAL RESEARCH ARTICLE

Application of Machine Learning Techniques for Predicting Hypertension Status and Indicators

Ibrahim Sule Haruna^{1*}, Sani Salihu Abubakar¹, Auwalu Ibrahim¹, Abdulhameed Ado Osi¹
and Usman Abubakar²

¹Department of Statistics, Aliko Dangote University of Science and Technology, 3244 Wudil, Kano State, Nigeria

²Department of Statistics, Jigawa State Polytechnic Dutse, Jigawa, Nigeria

ABSTRACT

One significant reason of the suffering of productive age around the world is hypertension, a wonderful illness that intentionally aggravates the symptoms of renal, brain, heart, and other ailments. Five machine learning approaches were used to classify the data according to training and testing sets: RF, CART, RT, SVM, and ANN. A confusion matrix and a receiver operating characteristic curve were employed to evaluate the efficiency of the models. This investigation assessed the effectiveness of five machine learning algorithms for forecasting hypertension and looked into the frequency of the condition. The results showed that 60.42% of the studied population suffered from hypertension. Furthermore, the comparison of machine learning models revealed that the artificial neural network outperformed the others, achieving AUC of 0.8694. The variable importance ratings highlighted diabetes and parental hypertension, which were the most significant predictors of hypertension. These findings can inform the development of effective predictive models and intervention strategies for hypertension management in Jigawa State of Nigeria.

INTRODUCTION

Hypertension is a great health issue that purposefully exacerbates the symptoms of kidney, heart, among others (Abubakar *et al.*, 2023). According to Mills *et al.*, (2020), 31.1% of adults worldwide (1.39 billion) were anticipated to have hypertension in 2010. According to Singh *et al.*, (2017), 1.56 billion adults are expected to have hypertension by 2025, with low- and middle-income nations continuing to have the majority of cases. It has been found that hypertension accounts for roughly 13.5% of deaths worldwide each year. Moreover, hypertension is directly responsible for 54% of all stroke artery diseases and 47% of all coronary heart diseases worldwide (Wang *et al.*, 2014). According to global data on responsibilities, risk factors affected 41% of all years of life with a handicap, according to the 2015 Infection, Harm, and Risk component research (Forouzanfar *et al.*, 2017).

To lessen this growingly questionable accountability in lower- and middle-class nations, like Africa, the World Health Organization, interested parties, and public health professionals declared non-communicable diseases a global distinction in 2011, as recommended at the UN meeting. As suggested by Adeloye *et al.*, (2015), there are about 17 million cardiovascular disease sufferers

worldwide, with these conditions causing about 7.5 million pains from hypertension and 57 million years of life lost due to incapacity, respectively, and accounting for about 12.8 (Singh *et al.*, 2017) and 3.7% of all death and years lost due to incapacity worldwide.

At the organizational, societal, and physician levels, controlling hypertension in Nigeria is fraught with difficulties. These difficulties are largely uncredited for the observed rise in the prevalence of issues among people with high blood pressure, despite the year's high progress of systolic remedies (Nelson, 2021). According to the Nigerian National Non-Communicable Diseases Survey Committee, 11.2% of men and women had hypertension in 1997, which at the time was responsible for 4.33 million cases in people over the age of 15 (Adeloye *et al.*, 2015). The lack of attention and poor management of hypertension in Nigerian society is one of the difficult factors affecting how this task is completed. (Kaima *et al.*, 2015). As a result, those who have hypertension frequently experience the health effects of cardiovascular issues such as ischemic cardiac disease, strokes, and heart failure (Asemu *et al.*, 2021).

Correspondence: Ibrahim Sule Haruna. Department of Statistics, Aliko Dangote University of Science and Technology, 3244 Wudil, Kano State, Nigeria. ✉ ibrahimsulebkd201@gmail.com.

How to cite: Sule, I. H., Salihu, S. A., Ibrahim, A., Ado, A. O., & Abubakar, U. (2025). Application of Machine Learning Techniques for Predicting Hypertension Status and Indicators. *UMYU Scientifica*, 4(2), 226 – 233. <https://doi.org/10.56919/usci.2542.023>

ARTICLE HISTORY

Received March 07, 2025

Accepted June 26, 2025

Published June 30, 2025

KEYWORDS

Machine learning,
Hypertension, Prevalence,
Prediction, Performance
evaluation



© The Author(s). This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License [creativecommons.org](https://creativecommons.org/licenses/by-nc/4.0/).

Additionally, machine learning (ML) can confirm indicator or a combination of indicators is most practical for forecasting hypertension (Alkaabi *et al*, 2020). Machine learning used in prediction models enables them to evaluate and comprehend clinical data that have been found as predictors for hypertension or not, in addition to additional variables like daily lifestyle and other biological markers. Predictive models are continuously updated as preventative measures designed to tailor the intensity of preventative efforts to people who have a higher chance of getting hypertension. Additionally, they aid in risk communication, which renders it simpler to recognize and choose those individuals who have a significant chance to grow up hypertension for therapeutic care. Finally, they help with resource allocation for potential hypertension burdens. Therefore, the objectives of the research are to;

- Find out how common hypertension is among Jigawa State patients,
- apply classification techniques for the classification of data associated with hypertension,
- Evaluate the performance of the techniques in (ii),
- Rate the importance of the variables that are most significantly contributing in the progress of hypertension.

MATERIAL AND METHODS

This section explained the study area and population, the description of data, and the methods of data collection. It also explained the classification techniques used in conducting the analysis using five ML techniques employed to classify the data based on training sets and performance based on testing sets, as well as the method to evaluate the performance of the models.

Study Population

Purposive sampling was used to investigate the information of 480 patients at selected Hospitals in Jigawa Central. The study subjects are made up of one hundred and ninety (190) Non-hypertension patients and two hundred and ninety (290) confirmed hypertension patients from the record department at three hospitals in Jigawa State.

Method of Data Collection

A cross-sectional study on the existing information of 480 patients was carried out at some selected hospitals in Jigawa Central, such as Dutse General Hospital, Birnin Kudu General Hospital, and Rasheed Shekoni Teaching Hospital in Jigawa State, Nigeria.

CLASSIFICATION TECHNIQUES

In this section, five classification techniques were explained for the prediction models of hypertension based on the data available for 480 patients.

Logistic Regression (LR)

LR fits data to a logistic curve to determine the probability of an event by analyzing the connection between several independent factors and a categorical dependent variable. When there are dichotomous dependent variables and continuous ones, binary LR is frequently employed (Hosmer *et al*, 2013). With the data prepared, logistic regression is applied to model the probability that a patient has hypertension. The logistic function maps the linear combination of input features to a value between 0 and 1, representing the predicted probability of hypertension. A threshold is then used to classify whether the individual is likely hypertensive or not. The model is trained using the training dataset, during which it estimates the coefficients for each predictor variable. These coefficients indicate the strength and direction of the relationship between each feature and the likelihood of hypertension. The equation represents logistic regression:

$$y = \frac{e^{(b_0+b_1x)}}{1+e^{(b_0+b_1x)}} \tag{3.1}$$

Classification and regression tree (CART)

A decision tree is a group of decision nodes that span from the root node to the leaf nodes and are joined by branches. Every decision node tests attributes, and every potential result produces a branch (Sandri and Zuccolotto, 2010). In this research, the CART model constructs a decision tree by recursively splitting the data into subsets based on the values of input features. CART builds a classification tree for hypertension prediction in this research, which is a binary classification task (i.e., hypertensive and non-hypertensive). The algorithm selects the feature and threshold that best separates the data into two groups at each node using an impurity measure. The tree continues to split until a stop, and the final tree consists of decision rules that are interpreted as result, then predict hypertension. After the model is built, it is tested on the test data, and performance is evaluated using accuracy, sensitivity, specificity, and the area under the ROC curve.

Support Vector Machines (SVMs)

The model created using SVR is based on structural risk maximization and solely depends on the subset of training data (Pisner *et al*, 2020). The SVM algorithm is trained on the labeled training data. The SVM finds the optimal hyperplane that separates the two classes (hypertensive and non-hypertensive) with the maximum margin. Then, SVM uses a kernel trick of radial basis function (RBF) to transform the data into a higher-dimensional space where a linear separator can be found. During training, the algorithm identifies a subset of training samples, known as support vectors, that are most critical in defining the boundary between classes. These points lie closest to the separating hyperplane and directly influence its position and orientation. After the model is trained, it is evaluated on the test dataset, and the performance is measured using metrics. The distance of the separation between a data point z_i and the decision boundary can be calculated as:

$$dj = \frac{WTz + b}{||W||} \tag{3.2}$$

Artificial Neural Network (ANN)

The ANN is a popular analytical method for resolving complicated problems (Bekesiene *et al*, 2021). The ANN model is designed with an input layer that accepts all predictor variables, one hidden layer with neurons that process the inputs using activation functions, and an output layer with a single neuron and a sigmoid activation function to predict the probability of hypertension. The model uses backpropagation to minimize a loss function, typically binary cross-entropy, by adjusting the weights through gradient descent. During training, the ANN learns patterns and associations between the input features and the target outcome (hypertension status). The output axon is represented by the signal $X(z)$, which is the outcome of the activation function $f(x)$ using the net total:

$$X(z) = g(\sum_{j=1}^D W_j Z_j) \tag{3.3}$$

Random Forest (RF)

RF is an ensemble learning method that uses categorization decision trees. It combines various decision trees to provide a final classifier. By building an ensemble of numerous uncorrelated decision trees and then averaging the results (Franklin, 2005). In this research, the data undergoes preprocessing to ensure quality and consistency. This involves handling missing values, normalizing or scaling numerical features, and encoding categorical variables into a format suitable for machine learning models. The dataset is then split into training and testing subsets. With the data prepared, the Random Forest model is trained on the training set. The algorithm builds an ensemble of decision trees using random samples and random feature subsets. Each tree votes on the outcome, and the majority vote determines the final prediction in classification tasks, or the average is taken in regression tasks. After the model is trained, it is applied to the test data to make predictions about whether patients are likely to have hypertension.

The Random Forest Algorithm in this case of classification performs as follows;

$$D(z) = g_0(z) + g_1(z) + g_2(z) + g_3(z) + \dots \tag{3.4}$$

Where the final model g is the sum of simple base models, g_i

DATA PARTITIONING

In this research, the hypertension-related data was partitioned into two parts (i.e., training and testing). The construction of both techniques was based on the input layer comprising all the indicators (depending on the number of hypertension indicators available under study), one hidden layer using a sigmoid activation function, and an output layer that displayed the classification result. In this research work, dataset partition rates for testing and training were allocated at 30% for testing the performance and 70% for training the model. The proposed methods created a model through comparing the association between the explanatory (Hypertension data) and study variable (Hypertension status) variables. After this

process, models validated the data by examining the predicted outputs, and as this result, both models are expected to perform better on hypertension datasets.

PERFORMANCE EVALUATION

In machine learning, it is essential to compare several approaches to classification statistically across numerous data sets (Pisner *et al*, 2020). Even though a variety of parameters can be employed to estimate the extent of classification accuracy for both fitting and validation objectives, these factors are inherently linked to the existence of imperfections in the results, even though the significance of these errors can vary depending on the classification goals. All models need to be tested using several assessment parameters before establishing a classification model (Xie *et al*, 2018). AUC/ROC, accuracy, reliability, precision, and recall are popular metrics to assess the efficacy of the models.

Confusion Matrix

A common method of evaluating the performance of the techniques is the confusion matrix (Abubakar *et al*, 2023). In this research, the confusion matrix was used to recognize correctly and wrongly classified data in the testing samples in both techniques, as shown in the following table and equations.

Table 1: Confusion Matrix

	Positive (1)	Negative (0)
Positive (1)	True Positive (TP)	False Positive (FP)
Negative (0)	False Negative (FN)	True Negative (TN)

Considering Table 1, the non-error rate (NER) and an error rate can be defined as follows.

$$Accuracy = \frac{TP+TN}{FN+TP+FP+TN} \tag{3.5}$$

Model accuracy can easily be examined from the ER and NER above. The sensitivity is equal to 1 for the correctly classified members, describes the model’s ability to correctly recognize objects belonging to the class, while the specificity is the opposite (Abubakar *et al*, 2024), and is defined as;

$$Sensitivity = \frac{TP}{TP+FP} \tag{3.6}$$

$$Specificity = \frac{FP}{FP+TN} \tag{3.7}$$

RESULTS

This section demonstrates the results obtained from the ML techniques, including classification results and performance evaluation metrics, including accuracy, sensitivity, specificity, and area under the curve using the receiver operation characteristics curve. R-Package was used throughout the work.

Prevalence of hypertension

The frequency of hypertension illustrated in Table 2 below shows the frequency of each category of variable.

In each category, a chi-square test of independence was conducted to determine the dependency of hypertension on each variable. The prevalence (60.42%) shows that male are more infected with hypertension than female; those that are self-employed followed by unemployed are more infected than employed, and retired; those that are married are often highly infected followed by divorce; those without diabetes had hypertension more than those

with diabetes; those that their parents have hypertension are more likely to be infected with hypertension. This highlights the likelihood of acquiring hypertension based on demographic characteristics. The p-value of 0.00 from chi-square test of independence throughout the variables indicated that there is dependency between each variable and hypertension status.

Table 2: Prevalence of hypertension

Variables	Category	Non-hypertension	Hypertension	P-value
Prevalence		39.58%	60.42%	
Gender				
	Male	0	235	
	Female	190	55	0.00
Marital status				
	Divorce	0	60	
	Married	59	274	0.00
	Single	111	0	
	Widow	20	0	
Employment				
	Employed	68	0	
	Retired	42	0	0.00
	Self-employed	0	238	
	Unemployed	80	52	
Diabetes				
	Diabetes	108	0	
	Non-diabetes	82	290	0.00
Life style				
	Light	59	115	
	Moderate	0	175	0.00
	Sedentary	131	0	
Exercise				
	Light	106	0	
	Moderate	84	119	0.00
	Sedentary	0	171	
Education				
	Primary	59	0	
	Tertiary	74	117	0.00
	Secondary	0	173	
	Uneducated	60	0	
Parental History				
	Hyper.	0	262	
	Non-hyper.	190	28	0.00
Locality				
	Rural	0	219	
	Urban	190	71	0.00
Age				
	0-30	0	25	
	31-60	90	256	0.00
	Above 60	0	160	

Logistic Regression

The logistic regression classifier performed slightly better than the other classifying correctly classified 35 out of 45 as hypertensive and misclassified 10. It also correctly classified 81 out of 104 and misclassified 23 which gives an accuracy of 77.85%, sensitivity of 10.99%, specificity of 39.66%, and AUC of 86.42% as shows in Table 4. These results with a small value of sensitivity indicated that the model failed to classify most of the data correctly. Hence performed badly.

Table 4: Performance metrics for logistic regression classifier

	H.	Non-H	A	S	SP	AUC
H	35	10	0.7785	0.1099	0.3966	0.8642
Non-H	23	81				

Key: H = Hypertensive; Non-H = Non-Hypertensive; A = Accuracy; S = Sensitivity; SP = Specificity

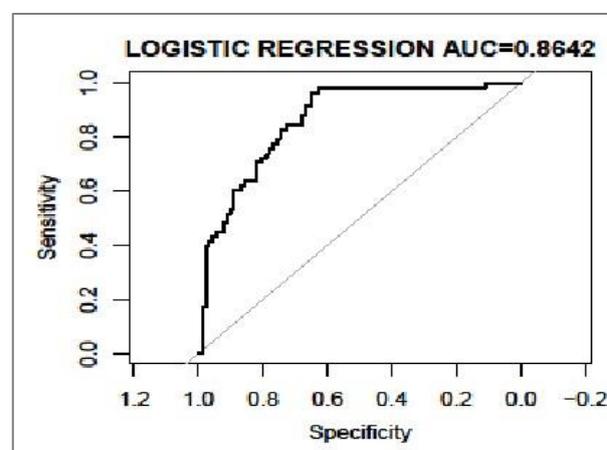


Figure 3: Receiver operation characteristics curve for LR

The receiver operation characteristics (ROC) curve in Figure 3 indicated the performance of the logistics regression classifier with AUC of 0.8694.

Artificial neural network

As a classifier for pattern recognition, the neural network performs better than LR and RF in ranking. This is possibly as a result that a neural network supports large data to perform a classification task. The neural network in Table 5 was classified 77 out of 97 as hypertensive and 38 out of 52 as non-hypertensive with an accuracy of 78.18%, sensitivity of 84.62%, specificity of 65.22%, and AUC of 86.94%. This classification results indicated that the model classified most of the data correctly.

Table 5: Performance metrics for artificial neural network classifier

	H.	Non-H	A	S	SP	AUC
H	77	20	0.7718	0.8462	0.6522	0.8694
Non-H	14	38				

Key: H = Hypertensive; Non-H = Non-Hypertensive; A = Accuracy; S = Sensitivity; SP = Specificity

CLASSIFICATION RESULTS

The data consist of 10 variables for 480 observations obtained from selected Hospitals in Jigawa Central like; Dutse General Hospital, Birnin-kudu General Hospital, and Rasheed Shekoni Teaching Hospital. This variable, hypertension-status, is categorical level which is whether a person has hypertension or not. So as to classify the 480 observations into hypertension and non-hypertension patients by using the Classification techniques on the label data. The data is made up of 480 entries, which is considered sufficient data to get the accurate Classifications task.

Random forest

The results obtained from the random forest illustrated how it classified the data with excellent performance. The result from random forest classifier in Table 3 below shows that it correctly classified 43 out of 59 as hypertensive and misclassified 16 and also correctly classified 75 out of 90 as non-hypertensive and misclassified 15 which gives an accuracy of 79.19%, 74.14% sensitivity, 82.42% specificity, and AUC of 78.11 for testing sample. This is possibly because the RF is an ensemble classifier that combines the predictions of several decision trees. The ROC curve in Figure 2 shows higher sensitivity, and this means that the model classified most of the data correctly with a small percentage of misclassification.

Table 3: Performance metrics for random forest classifier

	H.	Non-H	A	S	SP	AUC
H	43	16	0.7919	0.7414	0.8242	0.7811
Non-H	15	75				

Key: H = Hypertensive; Non-H = Non-Hypertensive; A = Accuracy; S = Sensitivity; SP = Specificity

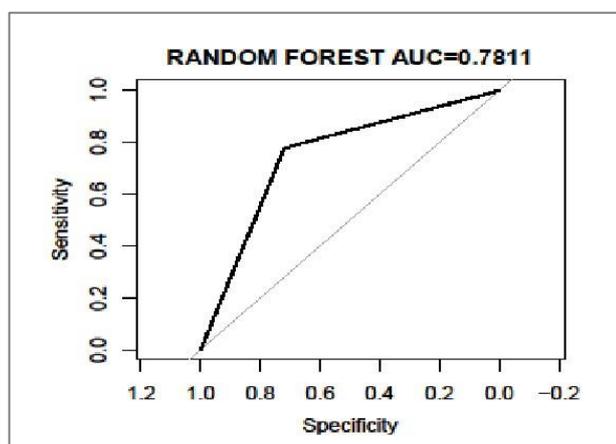


Figure 2: Receiver operation characteristics curve for RF

The receiver operation characteristics (ROC) curve in Figure 2 shows the highest sensitivity, indicating a very good performance of the classifier with AUC of 0.7811.

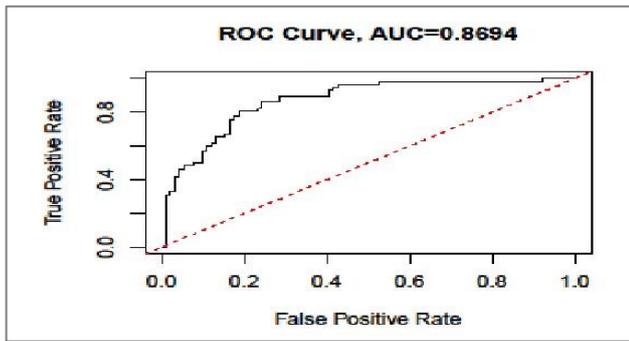


Figure 4: Receiver operation characteristics curve for ANN

Classification and regression tree

However, the result of classification and regression tree method in Table 6 below shows that it correctly classified 43 out of 59 as hypertensive and misclassified 16, it also classified 75 out of 90 as non-hypertensive and misclassified 15 which yields an accuracy of 79.19%, 74.14% sensitivity, 82.42% specificity, and 78.11% AUC respectively. In this case, the model is similar to the random forest classifier when compared to logistic regression and artificial neural networks.

Table 6: Performance metrics for CART classifier

	H.	Non-H	A	S	SP	AUC
H	43	16	0.7919	0.7414	0.8242	0.7811
Non-H	15	75				

Key: H = Hypertensive; Non-H = Non-Hypertensive; A = Accuracy; S = Sensitivity; SP = Specificity

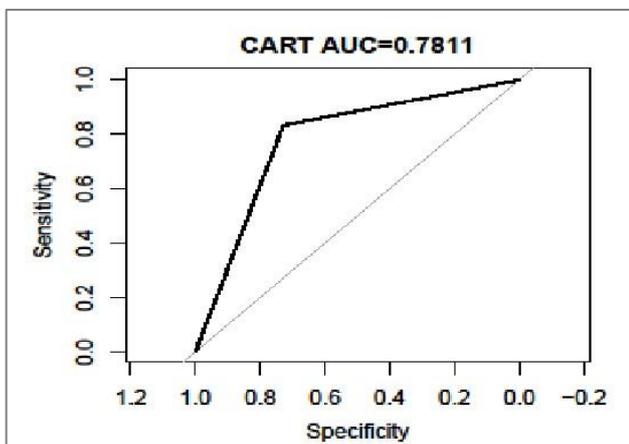


Figure 5: Receiver operation characteristics curve for CART

Figure 5 above shows the percentage of correctness classification made by CART classifier with AUC of 0.7811, similar to that of the random forest.

Support vector machines

Like other classification techniques, Table 7 shows the SVM performance for the classification task. The results indicated that it correctly classified 35 out of 45 as hypertensive and misclassified 10, it also classified 81 out of 104 as non-hypertensive which gives an accuracy of 77.85%, sensitivity of 60.34%, specificity of 39.66%, and

AUC of 77.83%, which indicated that both RF and SVM perform similarly in terms of accuracy but differ in sensitivity.

Table 7: Performance metrics for SVMs classifier

	H.	Non-H	A	S	SP	AUC
H	35	10	0.7785	0.6034	0.3966	0.7783
Non-H	23	81				

Key: H = Hypertensive; Non-H = Non-Hypertensive; A = Accuracy; S = Sensitivity; SP = Specificity

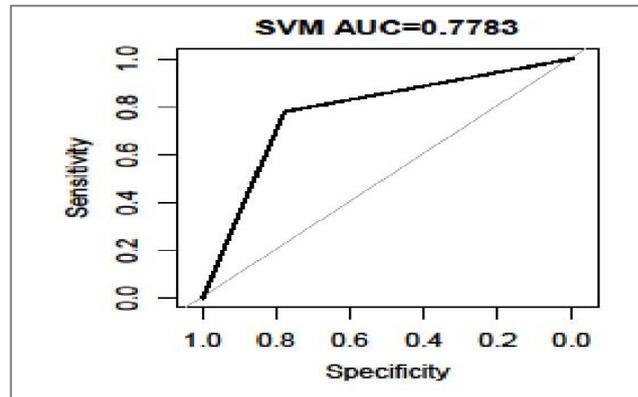


Figure 6: Receiver operation characteristics curve for SVMs

Performance evaluation and comparison of the techniques

The analysis demonstrated results from five techniques: RF, ANN, LR, CART, and SVM. The results illustrated in Table 8 compared the performance of the techniques based on accuracy, sensitivity, specificity, and area under curve respectively. The results show that the ANN better distinguished between a person with hypertension and not. The accuracy sometimes provided misleading information or results due to imbalance of data as it provided the overall performance of the techniques. Therefore, based on AUC, ANN (AUC=0.8694) performed better than other existing RF (AUC=0.7811), LR (AUC=0.8642), and CART (AUC=0.7811), though the additional SVMs (AUC=0.7783) performed poorly.

Table 8: Comparative evaluation of models performance

Models	Accuracy	Sensitivity	Specificity	AUC
RF	0.7919	0.7414	0.8242	0.7811
CART	0.7919	0.7414	0.8242	0.7811
LR	0.7785	0.1099	0.3966	0.8642
SVMs	0.7785	0.6034	0.3966	0.7783
ANN	0.7718	0.8462	0.6522	0.8694

Variable importance

The relative importance of the variable in the analysis illustrated in Table 9 and Figure 7 shows the contribution of the variable negatively or positively in the progression of hypertension. The variables with negative values are contributing negatively and imply that diabetes with a

value of -0.7158 is highly contributing to the progress of hypertension, followed by gender with a value of -0.2127, then exercise with a value of -0.17034. The variables with positive values implied no effect on the progress of hypertension, with 0.000 for locality, indicating that it has no effect.

Table 9: Relative importance of variables

S/N	Relative importance	Variable Names
1	-0.71583029	Diabetics.Status
2	-0.21273267	Sex
3	-0.17034692	Exercise
4	-0.05719819	Lifestyle
5	-0.05697489	Educ.qualification
6	0.00000000	Localty
7	0.02094819	Age
8	0.03468210	Employment.Status
9	0.22103162	Marital.Status
10	1.00000000	Parent.Hypertensive

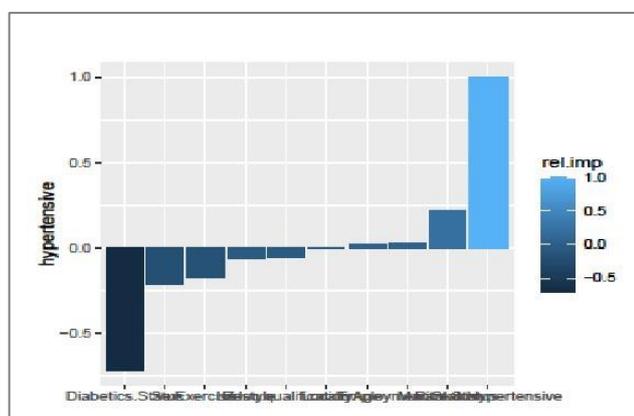


Figure 7: Variable importance chart

DISCUSSION

The analysis performed demonstrated the use of ML techniques in classifying patients as having hypertension or not. The analysis considers 10 hypertension-related variables as an indicators; these are age, gender, marital status, employment status, diabetes status, hypertension status, lifestyle, exercise, educational qualification, and parent hypertension status. The data consists of 480 observations, which are either hypertensive or non-hypertensive, and served as a dependent variable, while the other variables are the factors expected to contribute to the progress in high risk of high-risk hypertension; variables like age and weight are among the variables for boosting the infection. People who are non-diabetic and have a parental history are more susceptible to hypertension or high blood pressure. A purposive sampling technique is employed to select four hundred and eighty (480) samples in the overall prevalence of hypertension in the study, which is found to be 60.42 percent. Out of this percentage, those who are married, non-diabetic. Those whose parents are hypertensive and those between the ages of thirty (30) and sixty (60) are prevalent to hypertensive. According to the prevalence (60.42%), men are more likely than women to have

hypertension; self-employed people are more likely to have it than those who are employed or retired; married people are frequently highly infected and then divorced; people without diabetes had higher rates of hypertension than people with diabetes; and people whose parents have hypertension are more likely to have it themselves. This illustrates the risk of developing hypertension according to demographic traits. There is a dependency between each variable and the presence of hypertension, as evidenced by the chi-square test of independence's p-value of 0.00 for all the variables.

The ML techniques used in this research were very good on the classification task performed, with a higher percentage of correct classification and a very low percentage of misclassification. Similarly, the AUC was also used in this research to examine the model with high and low performance, which indicated that the model with a lower value of AUC outperformed better than other classifiers. The findings demonstrate that the ANN outperformed humans in determining whether a person had hypertension or not. Because accuracy provided the overall performance of the procedures, it occasionally produced inaccurate information or results owing to data imbalance. Thus, ANN (AUC=0.8694) outperformed the other existing RF (AUC=0.7811), LR (AUC=0.8642), and CART (AUC=0.7811) based on AUC, while the extra SVMs (AUC=0.7783) fared badly. Indicating that diabetes, with a value of -0.7158, is a major contributor to the progression of hypertension, followed by gender, with a value of -0.2127, and exercise, with a value of -0.17034, are the variables with negative values. The factors with positive values suggested no impact on the development of hypertension, while the location value of 0.000 indicated no effects at all.

CONCLUSION

This analysis investigated the prevalence of hypertension and compared the performance of five machine learning techniques in predicting hypertension. The results showed that 60.42% of the studied population suffered from hypertension. Furthermore, the comparison of machine learning models revealed that the artificial neural network outperformed the others, achieving AUC of 0.8694. The variable importance ratings highlighted diabetes and parental hypertension, which were the most significant predictors of hypertension. These findings can inform the development of effective predictive models and intervention strategies for hypertension management in Jigawa State of Nigeria.

REFERENCES

Abubakar, U., Osi, A. A., Muhammad, Y. I., Salisu, I. A., Muhammad, A. B., Muhammad, N., & Abubakar, W. (2024). Comparison of three distribution free classification techniques applied to crime data of Nigeria prior and post COVID-19 pandemic. *FUDMA Journal of Sciences*, 8(1), 345–353. Retrieved from fjs.fudutsinma.edu.ng [Crossref]

- Abubakar, U., Abubakar, A., Sulaiman, A., Ringim, H. I., Salisu, I. A., Osi, A. A., ... & Haruna, I. S. (2023). Application of artificial neural network for predicting hypertension status and indicators in Hadejia Metropolitan. *FUDMA Journal of Sciences*, 7(1), 284–289. [\[Crossref\]](#)
- Adeloye, D., Basquill, C., Aderemi, A. V., Thompson, J. Y., & Obi, F. A. (2015). An estimate of the prevalence of hypertension in Nigeria: A systematic review and meta-analysis. *Journal of Hypertension*, 33(2), 230–242. [\[Crossref\]](#)
- 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). [\[Crossref\]](#)
- Asemu, M. M., Yalew, A. W., Kabeta, N. D., & Mekonnen, D. (2021). Risk factors of hypertension among adults: A community-based study in Addis Ababa, Ethiopia. *PLOS ONE*, 16(4). [\[Crossref\]](#)
- Bekesiene, S., Smaliukiene, R., & Vaicaitiene, R. (2021). Using artificial neural networks in predicting the level of stress among military conscripts. *Mathematics*, 9(6). [\[Crossref\]](#)
- Forouzanfar, M. H., Liu, P., Roth, G. A., Ng, M., Biryukov, S., Marczak, L., Alexander, L., Estep, K., Abate, K. H., Akinyemiju, T. F., Ali, R., Alvis-Guzman, N., Azzopardi, P., Banerjee, A., Barnighausen, T., Basu, A., Bekele, T., Bennett, D. A., Biadgilign, S., ... Murray, C. J. L. (2017). Global burden of hypertension and systolic blood pressure of at least 110 to 115 mmHg, 1990–2015. *JAMA*, 317(2), 165–182. [\[Crossref\]](#)
- Franklin, J. (2005). The elements of statistical learning: Data mining, inference and prediction. *The Mathematical Intelligencer*, 27(2), 83–85. [\[Crossref\]](#)
- Hosmer, D. W., Jr., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression*. John Wiley & Sons. [\[Crossref\]](#)
- Kayima, J., Nankabirwa, J., Sinabulya, I., Nakibuuka, J., Zhu, X., Rahman, M., Highenecker, C. T., Katamba, A., Mayanja-Kizza, H., & Kanya, M. R. (2015). Determinants of hypertension in a young adult Ugandan population in epidemiological transition—The MEPI-CVD survey. *BMC Public Health*, 15(1). [\[Crossref\]](#)
- Mills, K. T., Stefanescu, A., & He, J. (2020). The global epidemiology of hypertension. *Nature Reviews Nephrology*, 16(4), 223–237. [\[Crossref\]](#)
- Nelson, I. O. (2021). Management of hypertension in Nigeria: The barriers and challenges. *Journal of Cardiology and Cardiovascular Medicine*, 6(1), 23–25. [\[Crossref\]](#)
- Sandri, M., & Zuccolotto, P. (2010). Analysis and correction of bias in total decrease in node impurity measures for tree-based algorithms. *Statistics and Computing*, 20, 393. [\[Crossref\]](#)
- Singh, S., Shankar, R., & Singh, G. P. (2017). Prevalence and associated risk factors of hypertension: A cross-sectional study in urban Varanasi. *International Journal of Hypertension*, 2017. [\[Crossref\]](#)
- Pisner, D. A., & Schnyer, D. M. (2020). Support vector machine. In *Machine learning* (pp. 101–121). Academic Press. [\[Crossref\]](#)
- Xie, Y., Zhu, C., Zhou, W., Li, Z., Liu, X., & Tu, M. (2018). Evaluation of machine learning methods for formation lithology identification: A comparison of tuning processes and model performances. *Journal of Petroleum Science and Engineering*, 160, 182–193. [\[Crossref\]](#)
- Wang, A., An, N., Xia, Y., Li, L., & Chen, G. (2014, September). A logistic regression and artificial neural network-based approach for chronic disease prediction: A case study of hypertension. In *2014 IEEE International Conference on Internet of Things (iThings), and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom)* (pp. 45–52). IEEE. [\[Crossref\]](#)
- Zhang, G., Eddy Patuwo, B., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14(1), 35–62. [\[Crossref\]](#)