

Machine Learning-Based Characteristic Survey of High Blood Pressure in Young Adults

Nne Rena Saturday¹; Igbudu Ezeunwo Kingsley²; Anthony Vivian Onyinyechi³

¹Department of Computer Science Rivers State University

²Department of Computer Science Rivers State University

³Department of Computer Science University of Port Harcourt

¹ORCID NO: 0009-0006-4731-7568

Publication Date: 2026/10/09

Abstract: Amongst the six major constituent of cardiovascular ailment worldwide, hypertension (high blood pressure) is the third in rank. High blood pressure (HBP) serves as a preceding event for many other illnesses, its subtle way of affecting patients makes it important to monitor and find a way of early detection before it becomes a serious illness in the lives of patients. A growing percentage of young adults and teenagers are suffering from hypertension (HBP) in recent times and poses a problem of high mortality rate. This study is focused on predicting the occurrence of HBP in a young adult based on young adults' responses in order to assist in curbing the problem. A logistic predictive model (binary logistic regression) was used to predict the possibility of a young adult being hypertensive. The model has an accuracy of 75% and a precision of 83% which makes it safe to say that this model predicted the possibility of high blood pressure in young adults.

Keywords: High Blood Pressure, Young Adults, Machine Learning, Elevated Blood Pressure, Hypertension.

How to Cite: Nne Rena Saturday; Igbudu Ezeunwo Kingsley; Anthony Vivian Onyinyechi (2026) Machine Learning-Based Characteristic Survey of High Blood Pressure in Young Adults. *International Journal of Innovative Science and Research Technology*, 11(9), 3328-3334. <https://doi.org/10.38124/ijisrt/26sep1381>

I. INTRODUCTION

Globally, elevated blood pressure has been the major cause of premature deaths as well as cardiovascular conditions. Increased occurrence of high blood pressure is prevalent in low-income and average income nations (Mills et al, 2020). There is increased number of persons with damaged end-organ between the ages of 20 and 60 from Sub-Saharan Africa as a result of high blood pressure compared to high-income nations. Also, there is a little percentage of adolescents between the ages of 10-19 suffering from high blood pressure in the Sub-Saharan African (Chen et al, 2023). However, there is a significant growth of young adults and teenagers who suffer from high blood pressure due to the life and living of these young people. Nearly 1 in 4 adults aged 20 to 44 have high blood pressure and less than half of adults (42%) with hypertension are diagnosed and treated, an estimated (46%) of adults with hypertension do not know that they are hypertensive (WHO, 2023). More threatening is the occurrence of this condition in children which cannot not be diagnosed until adulthood or at its terminal level (Katamba et al. 2020).

High Blood Pressure is a medical condition that largely increases the risks of heart malfunctioning, brain, kidney, etc., diseases (Kim et al, 2023). It is a condition where the blood

vessels are under intense pressure continually; thus, increasing stress on the vessels (WHO, 2019). There are risk factors in young adults for high blood pressure and they are namely: Age, family history, alcohol intake, obesity, smoking of cigarette, gender, socioeconomic status, physical inactivity, birth weight, neoplasm, aorta coarctation, diet, medications, etc. (WHO, 2023). This has led to speculations that elevated blood pressure in adolescents eventually leads to hypertension in adulthood (Yang et al, 2019). Several research has been done and is still on-going trying to predict the occurrence of high blood pressure in children, adolescent and young adults. Some have tried with various machine learning algorithms; applying hybridization of these algorithms. This paper is based on using a simple machine learning algorithm to analyze some key characteristic factors that seem to play a major role in young adults suffering from high blood pressure. It is pertinent to investigate this issue from the basic level so as to handle it from the root cause if it is possible; considering the prevalent lifestyle of young people with regards to social media addictions, digital transformation of life' activities and the "get rich quick syndrome" in Africa. Devices that measure body composition (body minerals, body fat, protein, water) to manage health issues and challenges are on the rise which necessitate the analysis of the different body composition (Seo et al, 2024). It has high reliability and validity as compared to the analysis of other parameters.

II. RELATED REVIEWS

Ji et al. (2021) reported that, medically, it is quite difficult to predict hypertensive-related issues. They adopted machine learning technique and data mining to predict hypertension issues. GB DT-based feature selection can screen out hypertension problems. LightGBM-based hypertension complications prediction model. The accuracy, F1, and AUC of the prediction models are 0.9189, 0.8888, and 0.9233 after 10-fold cross-validation and comparison analysis, which is much superior to other machine learning models. The proposed method can reliably identify hypertension problems, so clinicians can take preventive actions to decrease their impact. However, a supervised machine learning model and a bigger variation of models could serve as a better predictor.

Chai et al. (2022) examined the use of anthropometric measurement in machine learning models; since it enhances the non-invasive predictive models. An imbalanced dataset of 2461 samples were used in the model and was balanced using Synthetic Minority Oversampling Technique (SMOTE) and random under-sampling. Eight attributes were used in the dataset; namely: age, C index, ethnicity, gender, height, location, parental hypertension, and waist circumference. The metrics used for evaluation were accuracy, precision, sensitivity, specificity, F1-score, misclassification rate, and AUC. No model consistently topped all seven-performance metrics. It was established that the choice of machine learning models determines the prediction behaviour and outcome.

Aparna and Babu (2020) worked on Blood Pressure prediction using Machine Learning algorithms. They specifically used unsupervised reinforced learning. A regression algorithm was used to predict Systolic Blood Pressure (SBP) using a dataset collected from sources. The created system may also be a fundamental machine learning algorithm for Systolic Blood Level Sign Prediction. To give more accurate information on the SBP forecast, the system's accuracy is frequently upgraded. To further improve prediction abilities, the anomalies relating to invalid and/or less accurate values are frequently deleted.

Shrivastava, Chakkaravarthy and Asif Shah (2023) proposed a new machine learning method for predicting systolic and diastolic blood pressure using clinical characteristics. This method is being proposed to overcome the challenge of required multiple sensors in the phonocardiogram and electrocardiogram cuff signals-based solution. They considered the following clinical characteristics – age, gender, use of alcohol, smoking habit, blood sugar, history of heart disease, cholesterol levels, weight. The model was trained and tested with four machine learning techniques – K-Nearest Neighbors (KNN), Logistic Regression, Decision Tree, Random Forest; using different ratios on the training, validation and testing in order to enhance the accuracy of the model. Evaluation of the algorithm's performance was carried out using accuracy, recall, precision, F1 scores.

Tahir et al (2026) published on predicting future risk of

hypertension in young adults using machine learning, lifestyle analysis and explainable AI. Their research work used machine learning to predict the risk of hypertension in adults within the ages of 18-30. Random forest was reported to predicted on the accuracy of 97%, and to improve transparent prediction, they used explainable AI, particularly Shapley Additive Explanations to reveal how individual factors supports the prediction risks. Their results showed that data-driven and interpretable approaches is able to present individualistic insights for early intervention or prevention.

Araujo-M et al (2025) in prediction of hypertension in pediatric population using machine learning and transfer learning; a multicentric analysis of SAYCARE study; analyzed data from children and adolescents in South America, using transferred knowledge from CATBOOST model trained on the children's data and adopted onto the adolescent' data sample. The reported results were Random Forest, XGBoost, and LightGBM achieved high accuracy (0.90) for children, with XGBoost and LightGBM demonstrating superior recall (0.50) and AUC-ROC (0.74). The model without transferred learning (TL) performed poorly, with accuracy and recall values remaining low and AUC-ROC ranging from 0.46 to 0.56. After applying TL, model performance improved significantly, with CatBoost achieving an AUC-ROC of 0.82, accuracy of 1.0, and recall of 0.18.

III. METHODOLOGY

Logistic regression which falls under the family of supervised machine learning models was adopted, this logit model is a statistical model mostly used for analysis on prediction as well as classifying the variables. It makes an estimate for the probability of an event to occur, for example, if a tumor is malignant or not malignant; based on a given dataset of independent variables. Since the outcome is a probability, the dependent variable is bounded between 0 and 1. Therefore, in this paper, we seek to determine the probability of a young adult suffer from high blood pressure. In logistic regression, a logit transformation is applied on the odds, that is, the probability of success divided by the probability of failure. It is also commonly referred to as the log odds, or the natural logarithm of odds, and the logistic function is given by:

$$\text{Logit}(\pi) = 1 / (1 + \exp(-\pi))$$

$$\text{Ln}(\pi / (1-\pi)) = \text{Beta}_0 + \text{Beta}_1 X_1 + \dots + \text{Beta}_k X_k$$

where:

Logit (π) = dependent or (response) variable X = independent variable

In this logistic regression equation, logit (π) is the dependent or (response) variable and x is the independent variable. The beta parameter, or coefficient, in this model is commonly estimated via maximum likelihood estimation (MLE). This method tests different values of beta through multiple iterations to optimize for the best fit of log odds. All of these iterations produce the log likelihood function, and

logistic regression seeks to maximize this function to find the best parameter estimate. Once the optimal coefficient (or coefficients if there is more than one independent variable) is found, the conditional probabilities for each observation can be calculated, logged, and summed together to yield a predicted probability. For binary classification, a probability less than .5 will predict 0 while a probability greater than 0 will predict 1.

➤ Binary Logistic Regression

In this approach, the response or dependent variable is dichotomous in nature, i.e. it has only two possible outcomes (e.g. 0 or 1). Some popular examples of its use include predicting if an e-mail is spam or not spam or if a tumor is malignant or not malignant. Within logistic regression, this is the most commonly used approach, and more generally, it is one of the most common classifiers for binary classification.

The flowchart in figure 1 shows the flow and direction of the prediction process; from data collection to prediction of high blood pressure. The proposed system architecture of the system is displayed in figure 2; showing the young adults in the range of 18-25 years old, their responses forming the dataset. The dataset is prepared using the data wrangling technique and processes with the target of best fit simulation. Thereafter, invoke the classification technique (binary logistic regression) and the prediction is done based on the parameters (independent variables). The result is either the likelihood of occurrence of high blood pressure or non-occurrence of high blood pressure as shown in figure 2. The dataset that comprises of the age, height, weight, Gender, etc. as shown in figure 3 and table 1 in its raw state; the first five rows is displayed and to ensure the system is interacting with the appropriate dataset. Note the columns 4th, 5th and 6th containing text which inappropriate for the classification process that deals with 1 and 0. Also, the columns 3rd and 4th will be used to compute the BMI (Body Mass Index) of every entry.

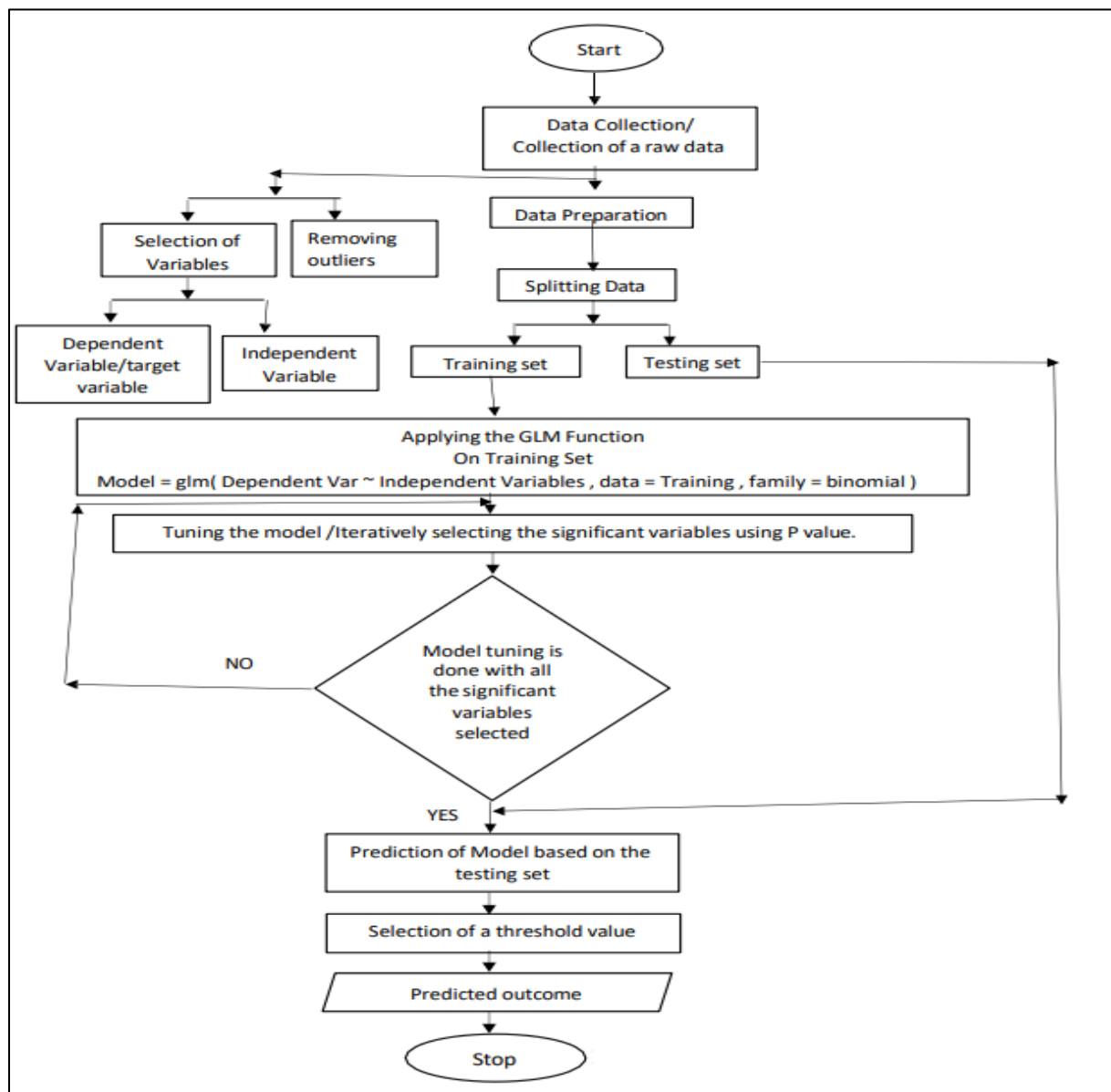


Fig 1 Flowchart of the Model Process

➤ *System Architecture of the Predictive System*

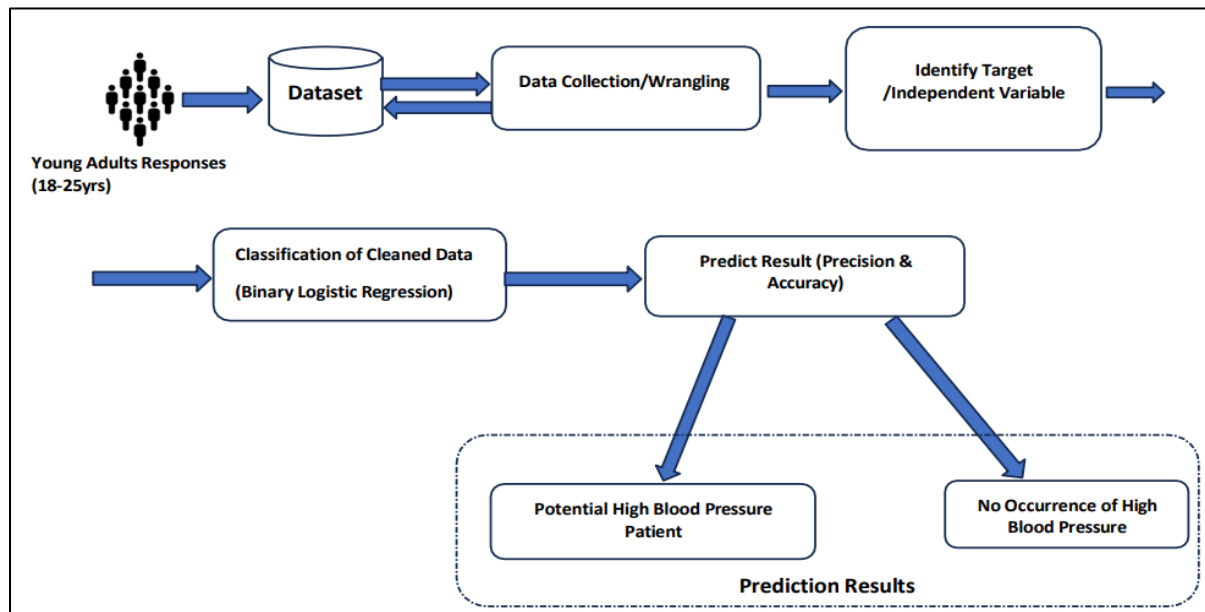


Fig 2 Proposed Architecture for High Blood Pressure Prediction

➤ *Data Evaluations and Analysis*

```

In [210]: df = pd.read_csv('datasett.csv')
          df.head(5)
    
```

	Age	Height	Weight	Gender	Smoking	Alcohol	Healthy eating	Workaholism	Getting angry	Life struggles	High blood pressure
0	20	163	48	female	never smoked	drink a lot	1	1	0	0	1
1	19	163	58	female	never smoked	drink a lot	1	1	1	0	1
2	21	176	67	female	tried smoking	drink a lot	1	1	1	1	1
3	22	172	59	female	former smoker	drink a lot	1	1	1	1	1
4	20	170	59	female	tried smoking	social drinker	1	1	0	0	1

Fig 3 First Five Records of the Dataset

```

In [202]: df['Smoking'].replace(['current smoker', 'never smoked', 'tried smoking', 'former smoker'], [1,0,1,0], inplace=True)
          df['Alcohol'].replace(['drink a lot', 'never', 'social drinker'], [1,0,1], inplace=True)
          df.head(5)
    
```

	Age	Height	Weight	Gender	Smoking	Alcohol	Healthy eating	Workaholism	Getting angry	Life struggles	High blood pressure
0	20	163	48	1	0	1	1	1	0	0	1
1	19	163	58	1	0	1	1	1	1	0	1
2	21	176	67	1	1	1	1	1	1	1	1
3	22	172	59	1	0	1	1	1	1	1	1
4	20	170	59	1	1	1	1	1	0	0	1

Fig 4 Rescaling of the Dataset

Table 1 Young People Survey Dataset

Age	Height	Weight	Gender	Ageing	Smoking	Alcohol	BMI	Family History	Healthy eating	Daily events	Itising worl	Workaholism	Thinking ahead	Hypochondria	Health	Mood swings	Appearance and gestures	Socializing
20	163	48	female	1	never smoked	drink a lot	18.06616734	Yes	4	2	2	4	2	1	1	3	4	3
19	163	58	female	3	never smoked	drink a lot	21.8299522	Yes	3	3	2	5	4	1	4	4	4	4
20	176	67	female	1	tried smoking	drink a lot	21.62984876	no	3	1	2	3	5	1	2	4	3	5
22	172	59	female	4	former smoker	drink a lot	19.94321255	yes	3	4	4	5	3	3	1	5	3	1
20	170	59	female	2	tried smoking	social drinker	20.41522491	no	4	3	1	3	5	1	3	2	3	3
20	186	77	male	1	never smoked	never	22.25690831	Yes	2	2	2	3	3	1	3	3	3	4
20	177	50	female	4	tried smoking	social drinker	15.95965399	no	4	3	5	5	3	1	3	5	4	5
19	184	90	male	1	current smoker	drink a lot	26.5831758	no	2	3	1	2	4	2	3	3	4	2
18	166	55	female	2	tried smoking	social drinker	19.95935549	no	1	1	2	4	2	2	4	3	4	4
19	174	60	female	3	never smoked	drink a lot	19.81767737	Yes	3	4	2	3	3	1	4	4	3	4
19	175	60	female	1	current smoker	social drinker	19.59183673	Yes	3	3	2	2	3	2	3	3	4	3
17	176	60	female	5	tried smoking	never	19.36983471	no	3	3	1	3	1	5	2	5	2	5
24	168	55	female	5	tried smoking	social drinker	19.48696145	Yes	3	4	3	4	5	5	5	4	4	2
19	165	55	female	2	never smoked	social drinker	20.2020202	Yes	3	5	3	2	3	1	3	3	5	3
22	175	57	female	5	current smoker	social drinker	18.6127449	no	1	4	5	2	5	5	5	4	4	1
18	177	77	male	1	former smoker	social drinker	24.57786715	yes	5	2		5	3	1	2	5	3	2
19	175	65	female	2	current smoker	drink a lot	21.2244898	no	3	2	3	5	5	1	3	3	4	3
20	168	65	female	5	tried smoking	social drinker	23.03004535	Yes	3	4	2	3	4	2	4	5	4	4
18	181	78	male	1	tried smoking	social drinker	21.80879705	no	4	2	4	4	5	1	3	3	4	3
18	188	90	male	2	former smoker	drink a lot	25.46401086	no	3	3	1	1	4	1	3	4	3	2
20	186	77	male	1	never smoked	never	22.25690831	no	2	3	2	3	2	1	3	3	3	4
24	186	85	male	1	tried smoking	social drinker	24.56931437	Yes	3	4	1	3	2	1	1	2	3	3
22	167	70	female	1	tried smoking	social drinker	25.0995016	Yes	3	2	3	2	3	2	3	3	3	3
20	168	54	female	3	tried smoking	social drinker	19.13265306	no	3	3	1	2	3	1	4	4	3	2
19	167	51	female	5	tried smoking	never	18.28677973	Yes	4	5	2	4	5	2	5	4	5	4
20	170	56	female	4	tried smoking	social drinker	13.37716263	Yes	3	3	3	2	4	1	3	4	4	2
22	165	63	female	2	tried smoking	social drinker	23.14049587	no	3	3	4	4	5	2	2	4	3	3
19	163	50	female	4	never smoked	social drinker	18.81892431	no	4	3	3	4	3	1	4	3	2	3
20	187	80	male	4	former smoker	social drinker	22.8774057	no	5	4	3	3	4	2	3	2	5	4
19	168	60	female	5	tried smoking	social drinker	21.2585034	no	3	4	5	5	5	2	4	4	5	4
25	183	87	male	4	former smoker	drink a lot	25.97867957	yes	3	1	2	1	3	1	4	3	4	3
22	177	80	female	1	tried smoking	social drinker	25.53544639	Yes	3	3	3	3	4	1	2	4	3	3
20	158	46	female	1	current smoker	social drinker	18.42653421	yes	3	3	1	1	2	1	1	2	2	4
27	170	70	female	1	tried smoking	social drinker	24.22145329	yes	4	5	2	3	1	2	2	3	3	3
20	177	67	male	1	tried smoking	social drinker	21.38593635	no	3	5	3	5	3	1	2	2	3	4

The dataset was rescaled into numerals so that the classification model can understand its input and compute the correct results as shown in figure 4.

IV. RESULTS AND DISCUSSIONS

```
In [218]: y_pred=logreg.predict(X_test)
print (X_test) #test dataset
print (y_pred) #predicted values
```

	Age	Height	Weight	Smoking	Alcohol	Healthy eating	Workaholism	\
15	18	177	77	0	1	1	1	
13	19	165	55	0	1	1	0	
17	20	168	65	1	1	1	1	
12	24	168	55	1	1	1	1	
14	23	175	57	1	1	0	0	
1	19	163	58	0	1	1	1	
18	18	181	78	1	1	1	1	
6	20	177	50	1	1	1	1	

	Getting angry	Life struggles
15	1	1
13	0	1
17	1	1
12	1	1
14	1	1
1	1	0
18	0	0
6	1	1

```
[1 1 1 1 0 1 1 0]
```

Fig 6 Prediction Using Logistic Regression Model

Accuracy: 0.75					
Recall: 0.8333333333333334					
Precision: 0.8333333333333334					
CL Report:		precision	recall	f1-score	support
0	0.50	0.50	0.50		2
1	0.83	0.83	0.83		6
accuracy			0.75		8
macro avg	0.67	0.67	0.67		8
weighted avg	0.75	0.75	0.75		8

Fig 7 Classification Report of the Predictive Model

The result is shown in figure 6, capturing the prediction for eight entries (8), as displayed horizontally. We have six 1s and two 0s; which means that six young adults are likely to have high blood pressure condition and two young adults will not have high blood pressure condition. Adopting the confusion matrix evaluation technique, we recorded 75%

accuracy and precision of 83% as shown in figure 7. The F1-score for each evaluation metric is recorded also in figure 7. It measured the class-wise performance of the model by combining precision and recall values which offers a trade-off. The recall recorded 83% of the positive class samples in the dataset identified correctly.

Predictor for High Blood Pressure

Please answer the following questions

How old are you?
18

What is your height?
177

What is your weight?
77

What is your gender?
☒ Male
☐ Female

Do you Smoke?
Former Smoker

Do you Drink?
Social Drinker

Do you Eat Healthy?
☒ Yes
☐ No

Do you Stress a lot?
☒ Yes
☐ No

Do you get angry easily?
☒ Yes
☐ No

Do you have life Struggles?
☒ Yes
☐ No

Predict

Result: You are likely to have high blood pressure!

Fig 8 User Interface of the Implemented Model

V. CONCLUSION AND FUTURE DIRECTION

The prediction of high blood pressure is quite complex as it is almost impossible from the medical perspective. Machine learning has offered a workable approach to make predictions of imminent circumstances. This paper offers a predictive model for high blood pressure in young adults using machine learning which has been designed and tested with the dataset of responses from young adults. Early diagnosis of possible high blood pressure can now be achieved from anywhere with little or no cost by leveraging this model. The model has an accuracy of 75% and a precision of 83% which makes it safe to say that this model predicted the likelihood of high blood pressure in young adults excellently. We have been able to implement this prediction providing a graphical user interface (GUI) using a very basic machine learning technique – binary logistic regression as a starting point. However, future research can be done with other techniques being implemented with a more robust and user-friendly interface.

REFERENCES

- [1.] Araujo-Moura K, Souza L, de Oliveira TA, Rocha MS, De Moraes ACF and Chiavegatto Filho A (2025) Prediction of Hypertension in the Pediatric Population Using Machine Learning and Transfer Learning: A Multicentric Analysis of the SAYCARE Study. *Int. J. Public Health* 70:1607944. doi: 10.3389/ijph.2025.1607944
- [2.] Chen A., Waite L., Mocumbi A.O., Chan YK., Beilby J., Ojji D.B., & Stewart S., (2023). Elevated Blood Pressure Among Adolescents in Sub-Saharan Africa: A Systematic Review and Meta-Analysis. *The LANCET Global Health* 11(8), 1238-1248.
- [3.] Katamba, G., Agaba, D. C., Migisha, R., Namaganda, A., Namayanja, R. & Turyakira,
- [4.] E. (2020). Prevalence of hypertension in relation to anthropometric indices among secondary adolescents in mbarara, southwestern uganda, *Italian Journal of Pediatrics* 46(1): 1–7.
- [5.] Kim, H.C., Lee, H., Lee, HH. *et al.* (2023). Korea Hypertension Fact Sheet 2022: Analysis of Nationwide Population-Based Data with a Special Focus on Hypertension in the Elderly. *Clin Hypertens* 29, 22. <https://doi.org/10.1186/s40885-023-00243-8>
- [6.] Mills, K.T., Stefanescu, A. & He, J. The global epidemiology of hypertension. *Nat Rev Nephrol* 16, 223–237 (2020). <https://doi.org/10.1038/s41581-019-0244-2>
- [7.] Seo, J.-W., Lee, S., & Yim, M. H. (2024). Machine Learning Approach for Predicting Hypertension Based on Body Composition in South Korean Adults. *Bioengineering*, 11(9), 921. <https://doi.org/10.3390/bioengineering11090921>
- [8.] Shrivastava A, Chakkaravarthy M & Asif Shah M, (2023). A New Machine Learning Method for predicting systolic and diastolic blood pressure using clinical characteristics. *Elsevier*, 4(100219), 2772-4425
- [9.] Tahir, N. Shafi, D. Shehzad, N. U. Haq and I. u. Din, "Predicting Future Risk of Hypertension in Young Adults Using Machine Learning, Lifestyle Analysis, and Explainable AI," 2026 6th International Conference on Electrical, Computer and Energy Technologies (ICECET), Rome, Italy, 2026, pp. 1-6, doi: 10.1109/ICECET65726.2026.11632679.
- [10.] World Health Organization, (2023). *A Report on Hypertension*. Retrieved from <https://www.who.int/news-room/fact-sheets/detail/hypertension>
- [11.] World Health Organization, (2023). World Health Organization Questions and Answer on Hypertension. Accessed by December, 2023
- [12.] World Health Organization, (2023). Hypertension. <https://www.who.int/news-room/fact-sheets/detail/hypertension> Accessed by December, 2023
- [13.] Yang L, Kelishadi R, Hong YM, Khadilkar A, Nawarycz T, Krzywinska-Wiewiorowska M, et al. (2019). Impact of the 2017 American Academy of Pediatrics Guideline on Hypertension Prevalence Compared with the Fourth Report in an International Cohort. *Hypertension*. 74(6), 1343–1348.