International Journal of Computer Applications (0975 – 8887)
Volume 186 – No.14, March 2024

Optimizing Hypertension Risk Classification through Machine Learning

Idongesit Umoh
Department of Internal Medicine,
University of Uyo, Nigeria

Victoria Essien
Department of Computer Science,
Akwa Ibom State
University, Nigeria

Saviour Inyang
Department of Computer Science,
Akwa Ibom State
University, Nigeria

ABSTRACT
Hypertension, a universal health concern worldwide, significantly contributes to cardiovascular disease and premature mortality. This study explores the classification of hypertension risk through machine learning techniques, aiming to enhance diagnosis support and preventive measures. Notable studies in this field have highlighted demographic trends, risk patterns, and prediction models using various algorithms. This study leverage on a Framework-Based Method, Where the data collection in this study involved observation and expert consultation, resulting in a dataset comprising 298 hypertensive patients' records. Data preprocessing was conducted using principal component analysis for feature selection, ensuring the relevance of variables like age, blood pressure, and lifestyle factors. Furthermore, the Classification processes employed Support Vector Machine (SVM), Decision Tree, and General Linear Model algorithms. Where SVM gave 90%, DT gave 83% and GLM gave 64% on accuracy. Performance evaluation metrics, including accuracy, sensitivity, and precision, were used to assess model efficacy. SVM emerged as the best-performing model and was deployed in a web-based interface for real-time hypertension risk classification. This study underscores the significance of machine learning in hypertension management, offering valuable insights into risk assessment and preventive strategies. The integration of SVM into a user-friendly interface enhances accessibility and empowers healthcare professionals and individuals to make informed decisions, ultimately mitigating the burden of hypertension-related complications.

Keywords
Hypertension, Machine Learning, Classification Risk

1. INTRODUCTION
Hypertension, commonly referred to as high blood pressure (HBP), is a prevalent and consequential health issue marked by increased blood pressure within the arteries. This condition stands as a primary contributor to cardiovascular disease and premature mortality on a global scale. The occurrence of hypertension is noteworthy, encompassing approximately 31.1% of adults (equivalent to 1.39 billion individuals) globally in 2010. The prevalence of adults grappling with hypertension surged from 594 million in 1975 to 1.13 billion in 2015, notably pronounced in low- and middle-income nations [1]. The escalating occurrence of hypertension, particularly in low- and middle-income nations, has been underscored by the World Health Organization (WHO). In response, the WHO has issued guidelines for the pharmacological management of hypertension in adults, aiming to bolster efforts in its prevention and control [2]. However, Elevated blood pressure, commonly known as hypertension, poses a substantial risk for a range of cardiovascular conditions such as coronary disease, left ventricular hypertrophy, valvular heart diseases, cardiac arrhythmias (such as atrial fibrillation), cerebral stroke, and renal failure [3]. Also, widespread of research have indicated that elevated blood pressure stands as a significant factor contributing to the risk of heart failure, atrial fibrillation, and chronic kidney disease [4]. In the context of population-attributable risk, hypertension emerges as the foremost modifiable risk element influencing the occurrence of premature cardiovascular disease. Preventing the age-related elevation in blood pressure and rigorously managing an established hypertension plays pivotal role in mitigating vascular repercussions, potentially eliminating a significant portion of cardiovascular disease. Again, physiologically normal blood pressure levels are significantly lower than traditionally defined, suggesting that cardiovascular disease is primarily induced by a rightward shift in the population distribution of blood pressure [5]. Nevertheless, Elevated blood pressure (EBP), along with cigarette smoking, diabetes mellitus, and lipid abnormalities, stands as a prominent modifiable risk factor for cardiovascular disease (CVD). Despite high prevalence, the evidence suggests that a biologically normal BP level in humans is significantly lower than traditionally used in clinical practices and research, potentially leading to an underestimation of BP's role as a risk factor for CVD. Hence, this study aims to classify the risk of hypertension through an examination of factors that can be modified as well as those that are unmodified factors that pose risk of hypertension. By gaining insights into these factors, targeted interventions can be developed to prevent and control hypertension, subsequently lowering the risk of associated complications. Modifiable factors, encompassing aspects like excessive salt intake, smoking, low physical activity, overweight, and excessive alcohol consumption, are subject to an individual's attitudes and behaviors. On the other hand, non-modifiable factors, including age, family history of hypertension, gender, and genetic makeup, remain beyond alteration. Concentrating on several risk factors has the potential to enhance public health and alleviate the global burden of hypertension, a significant contributor to cardiovascular disease and premature mortality.

2. RELATED LITERATURE
In recent years, there has been a surge in research utilizing machine learning to explore and classify hypertension, aiming to enhance diagnosis support. The following summarizes key findings from notable studies in this domain. [6] highlighted the challenges in managing hypertension due to its high prevalence, especially among certain demographic groups. Their cross-sectional studies examined changes in hypertension awareness, treatment, and control over time, forming the basis for preventive methods. [7] conducted a study in the Middle Eastern population, employing data mining to identify risk patterns associated with hypertension. Their prediction models, utilizing decision tree algorithms, demonstrated the influence of factors like systolic and diastolic blood pressure, age, and

waist size in predicting incident hypertension. [8] systematically characterized and predicted human hypertension genes, exploring distinguishing features of known hypertension genes. Their machine-learning method utilized these features to forecast novel hypertension genes. Also, [9] focused on predicting incident essential hypertension within the next year using electronic health records (EHRs) and the XGBoost machine learning algorithm. Leveraging EHR data, their study aimed to build a risk prediction model for timely detection of hypertension.

Nevertheless, [10] analyzed SPRINT trial data using machine learning to predict individual treatment results for intense blood pressure. Their findings emphasized the importance of considering heterogeneous treatment effects to estimate individual benefits accurately. Meanwhile, [11] utilized deep neural networks to predict blood pressure variability and mean blood pressure values. Their study aimed to address the challenge of reliably categorizing hypertension status using out-of-office blood pressure measures. [12] proposed an automated detection system for the severity of hypertension using electrocardiography (ECG) signals. Their study incorporated an optimal bi-orthogonal wavelet filter bank and machine learning to classify hypertension severity. [13] explored hypertension prediction models in the Qatar Biobank Study, comparing decision tree, random forest, and logistic regression algorithms. Their research demonstrated the potential of machine learning to develop non-invasive predictive models for hypertension screening. Furthermore,[14] delves into the persistent global impact of hypertension, particularly in Africa, emphasizing the need for advanced diagnostic and treatment approaches. Utilizing machine learning, the study proposes an innovative algorithmic method for intelligent hypertension classification, aiming to enhance cardiology services and early syndrome identification. Employing artificial intelligence, the research conducts exploratory data analysis to reveal clinically relevant evidence and optimize medical knowledge. The application of a Support Vector Machine for severity index classification yields an 83% accuracy rate, demonstrating its efficiency for hypertension classification and potential benefits for medical practitioners and healthcare departments. [15] introduces a computationally intelligent system for early detection and classification of cardiovascular diseases (CVD). Using clinical and ECG recordings from patients with hypertension at the University of Uyo Teaching Hospital, three machine learning models (Random Forest Ensemble, Support Vector Classifier, and Artificial Neural Network) were employed. Results indicate promising accuracy and precision scores for these models, supporting their potential in assessing and classifying early syndromic conditions of CVD. Nevertheless, these studies collectively contribute valuable insights into hypertension management and prediction using machine learning techniques. From uncovering risk patterns to predicting incident hypertension, these approaches hold promise in advancing the understanding and early detection of hypertension-related complications.

3. METHODOLOGY

In this research, a Framework Based Method (FBM) [16] is adopted which encapsulated all the process that is involve in the classification process of Hypertension using machine learning approach. Hence, we present the framework-based method in Figure 1. The framework presents all the different components or constituents that shows the different processes this research carried out from the point of data collection, Classification process and to the point of model deployment.

3.1 Data Collection

A. Collation

Collation of data is the process of gathering quantifiable data on certain factors with the goal of evaluating results or acquiring insights. Regardless of the subject of study, data collecting is usually the first and most significant phase in the research process. In this study, data was acquired through observation. The observation method is a technique for observing and describing an individual's behaviour. It is a method of gathering essential data and information through observation, as the name implies. This is also known as an application related since the researcher must build a connection with the responder and must do so by immersing himself or herself in their environment. Only then may he record and take notes using the observation approach. Hence, in this research, data was gathered over a period of one month in university of Uyo teaching hospital, where folders of different patients suffering or being diagnosis of hypertension in cardiology unit of the Hospital was access and data extracted.

B. Participants

The expression participant refers to the information collected from each study participant here in this research the individual participants are the patients, medical expert and the researcher. Individual participant data was collected in a cardiology department of the University of Uyo Teaching Hospital. The patient constitutes a cross-section of those who were diagnose of hypertension and were getting treatment from the University of Uyo for a minimum of 1 year and above. Also, a chief medical Expert in the unit of cardiologist in person of Dr. Umoh was present during the data gathering process to assist on issues and guide the process of data collection. Hence, table 1depicts study participants.

Table: 1 Study participants

Weeks	No of Patient Folder Accessed	No of Patient Folder Selected	Medical Expert on duty
Wk 1	78	25	*Dr. Umoh*
Wk 2	120	45	*Dr. Umoh*
Wk 3	65	17	*Dr. Umoh*
Wk 4	67	35	*Dr. Umoh*
Wk 5	98	23	*Dr. Umoh*
Wk 6	25	14	*Dr. Umoh*
Wk 7	31	19	*Dr. Umoh*
Wk. 8	145	21	*Dr. Umoh*
Wk. 9	82	34	*Dr. Umoh*
Wk. 10	18	12	*Dr. Umoh*
Wk. 12	27	19	*Dr.. Umoh*
Wk 13	31	34	*Dr. Umoh*

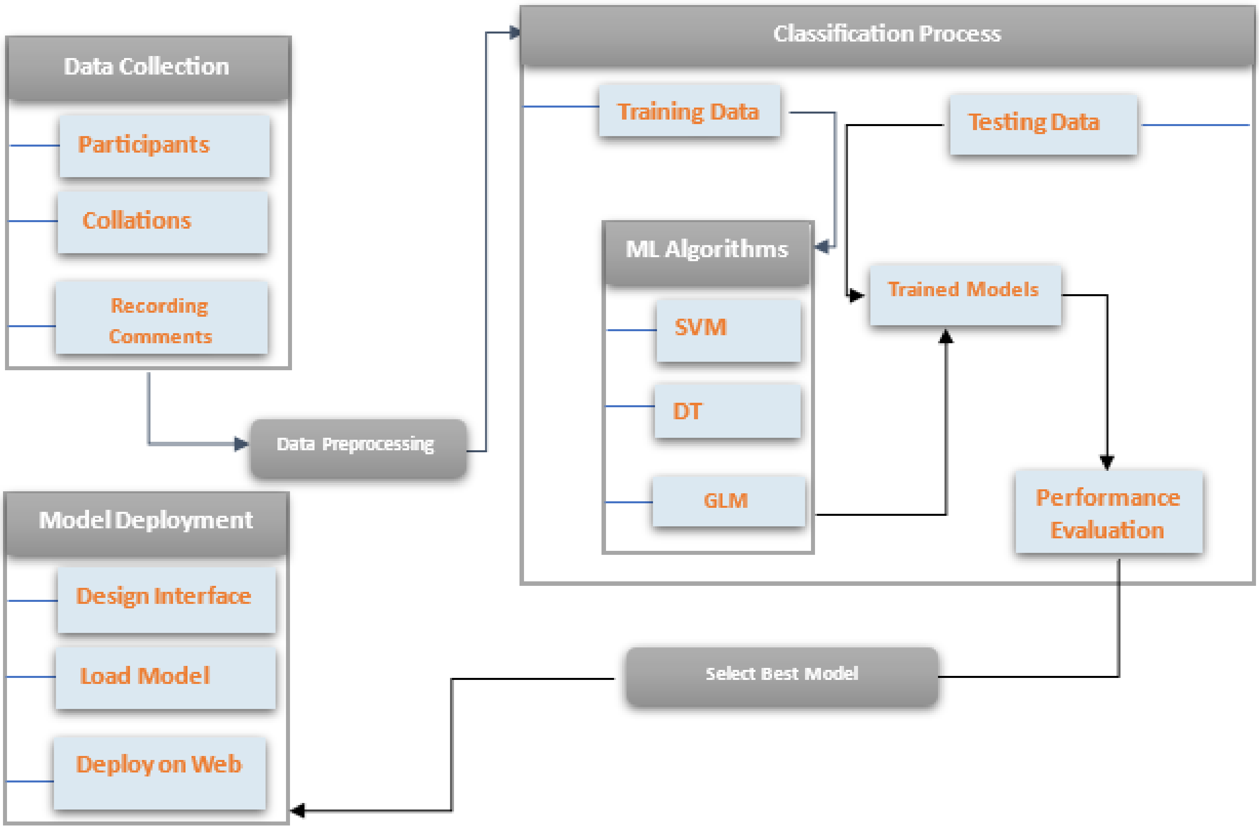


Figure 1: Machine Learning Framework for Classification of Hypertension

C. Recording Comments

The recording of the clinical data was done through the method of searching for each information in patient folders which was selected from the cardiology unit of the University of Uyo Teaching Hospital, a total of 787 hypertension patent folders were accessed and 298 were selected for data collection. Also, some factors influence our decision on the selection of the folders such as that the folder to be selected for the study must have been a diagnose of hypertension patient for at least one year and he or she must currently be undergoing treatment in the hospital also each folder must constitute all the parameters that was earlier categorized in table 2 in this research , this factor was a very uphill tasks in this research work because since folders are always move from the record units to cardiology section on every clinical days so it is always difficult to keep track of every information that should be or constitute in the parameter that we choose in this research.

3.2 Data Pre-Processing

Data preprocessing is an important step to prepare the data which will be used for the formation of a model. Data cleaning, data transformation, and feature selection are all key phases in data preprocessing. In this study, Principal component Analysis will be used in feature selection. Data cleaning and transformation are ways for removing outliers and standardizing data so that it may be utilized to build a model more readily. Variables, traits, fields, attributes, and dimensions are all terms used to describe features. Table2 shows the data attributes and their units in the dataset in this study.

Table:2 Data Attributes Categorization

Attributes	Category/units
Age	Numeric
Systolic Pressure (SP)	Numeric (mmHg)
Diastolic Pressure (DP)	Numeric (mmHg)
Family History (Hx)	Boolean (True or False)
Sodium Intake	Numeric (135-145 mmoI/I)
Cholesterol	Numeric (3.8-6.5 mmoI/I)
Blood Sugar Level (FBS)	(3.0-5.5 mmoI/I)
Alcohol Intake	Boolean (Yes or No)
Diabetic (DM)	Boolean (Yes or No)
Body Mass Index (BMI)	Numeric ($BMI = \frac{weight(kg)}{[height\ (m)]^2}$)
Risk	Boolean (Low or High)

3.3 Classification Process

A. Model Training and Testing

In machine learning, the training data set is the real dataset that is used to fit the classifier to execute different actions. That's the actual data that the models learn as part of the continual development process, using various APIs and algorithms to teach the machine to work autonomously. One of the tasks in machine learning is to research and develop algorithms that can learn and adapt from and predict data. These algorithms typically produce a computational formula from input data and producing information predictions or judgments. In this research work our training set is derived from a statistical field data which was collected from hospital, and undergone preprocessing stages in order to eliminate some noisy and redundant data. Our data was divided into a ratio of 0.7%(training) and 0.3% (test set) for the SVM classification process and for the prediction process which was done using Logistic regression where the split ratio was 0.7%(training) and 0.3%(testing).

B. Machine Learning Algorithms

The research on hypertension classification has adopted several Machine Learning algorithms, including Support Vector Machine (SVM), Decision Tree, and Generalized Linear Model. These algorithms are well-suited for classification tasks due to their distinct characteristics. SVM perform wells in managing datasets with numerous dimensions and proves especially advantageous when the data exhibits a distinct separation margin. Decision trees offer simplicity in interpretation and versatility in handling both numerical and categorical data, making them adept at pinpointing crucial features in the classification of hypertension. consequently, the Generalized Logistic Model serves as a regression analysis variant widely employed for binary classification, rendering it well-suited for predicting hypertension by considering various risk factors. Each algorithm exerts a distinct influence on hypertension classification. SVM, recognized for its adeptness with intricate data, skillfully discerns clear class boundaries, aiding in the identification of specific hypertension patterns. Decision trees shine in spotlighting pivotal features, unraveling the intricacies contributing to hypertension. Meanwhile, the Generalized Logistic Model, tailored for binary classification tasks, emerges as a fitting choice for predicting hypertension based on a diverse array of risk factors.

B. Performance Evaluation

The assessment of performance plays a pivotal role in the field of machine learning, encompassing a diverse set of metrics designed to gauge the accuracy and efficacy of an algorithm. In classification scenarios, two widely employed metrics are the confusion matrix and ROC graph. The confusion matrix delineates true positives, true negatives, false positives, and false negatives, offering a comprehensive view of model performance. Conversely, the ROC graph visually represents the true positive rate against the false positive rate. Classification tasks further utilize metrics like precision, recall, sensitivity, specificity, F-measure, and Matthew's correlation. In regression tasks, distinct metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared are commonly applied to assess model performance. The selection of a specific metric hinges on several factors, encompassing the research objectives, business goals, and the inherent strengths and limitations of each metric.

3.4 Model Deployment

This research will integrate the best performing model form their algorithm adopted in this research in to web application so that it functionalities can be deployed online and make it useful. Hence, integrating a machine learning model into a web app significantly improves accessibility for users unversed in machine learning. This incorporation allows individuals lacking expertise in the field to effortlessly interact with and harness the predictive capabilities of the model. The user-friendly interface simplifies the intricate classification process, enabling users to input pertinent data and obtain actionable insights without grappling with the complexities of the underlying algorithms. The Shiny app's intuitive design ensures a smooth and engaging user experience, rendering it a valuable instrument for those aiming to comprehend and utilize the classification model. This integration of machine learning empowers users across various sectors, granting them the ability to make informed decisions based on predictive outcomes. Beyond merely bridging the gap between machine learning sophistication and end-users, the app's deployment underscores the significance of democratizing technology, positioning it as an essential and practical tool for a diverse range of audiences.

4. RESULTS AND DISCUSSIONS

4.1 Data Structure

We present the data structure of the dataset we employe in this study in figure 2. From figure 2, the structure of the data used in this study consist of 90 training observations and 11 variables where the first 10 variables are independents variables that depends on the class variable which is the Risk variable that signifies the class variable that will be used in the classification process of the ML part of this study.

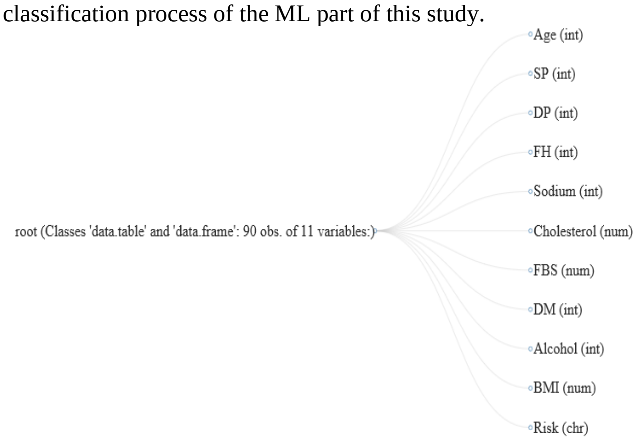


Figure 2: Data Structure

4.2 Correlation Analysis

We present the correlation analysis using correlation matrix in order to present the relationship level of each variable in the dataset in figure 3. So that more informed decision towards the deployment of ML model in the study.

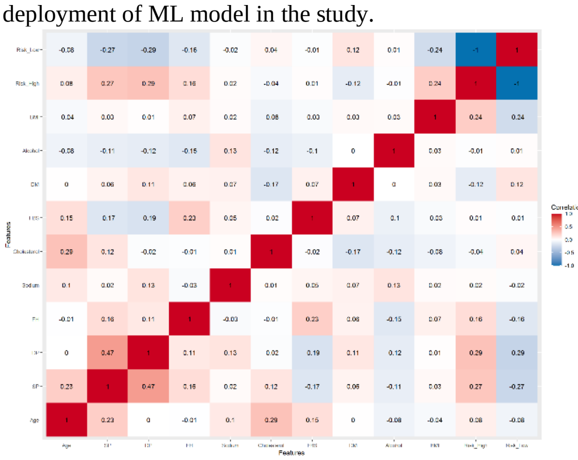


Figure 3: Correlation Matrix

4.3 Data Preprocessing

Data preprocessing plays a pivotal role in the realm of machine learning, encompassing tasks such as data cleansing, transformation, and priming data for analysis [17]. Within the domain of feature selection during data preprocessing, one significant methodology employed is Principal Component Analysis (PCA). PCA serves as a means to convert a set of interrelated variables into a condensed collection of uncorrelated variables termed principal components (PCs). Each PC represents a linear amalgamation of the original variables, weighed by their respective significance. Hence, we present the feature selection that was carried out on this research through the use of PCA to rank individual feature in the data sets as principal components which is presented in figure 4.

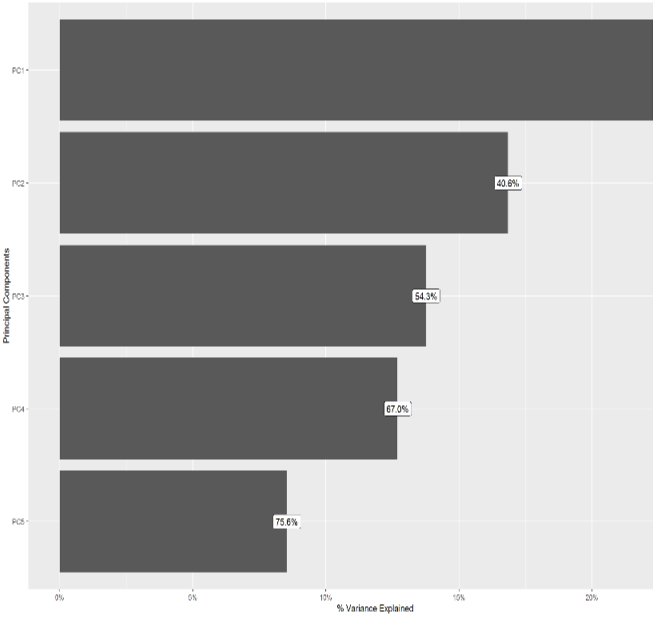


Figure 4: Data Structure

Furthermore, we present the features and their relative importance each in the dataset which is depicted in figure 5.

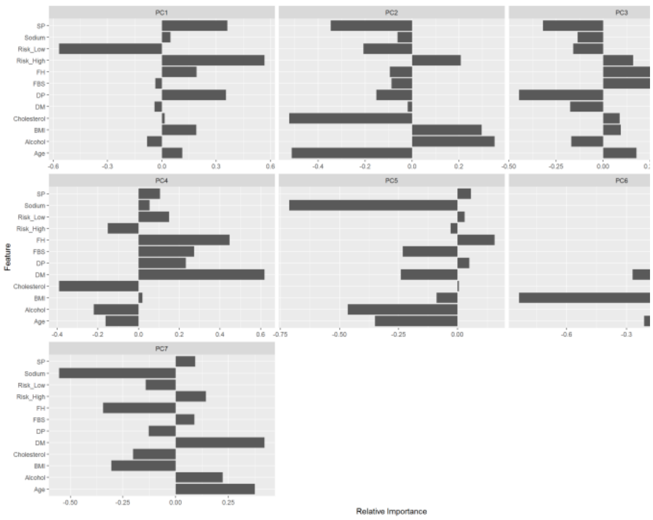


Figure 5: features & Relative importance

Furthermore, we present a Q-Q (Quantile-Quantile) plot that provides a visual comparison between the quantiles of the observed data and the quantiles of the expected distribution.
It allows the assess of a dataset to identify if it follows a particular theoretical distribution. Hence, the QQ pot for the data set is presented in figure 6.

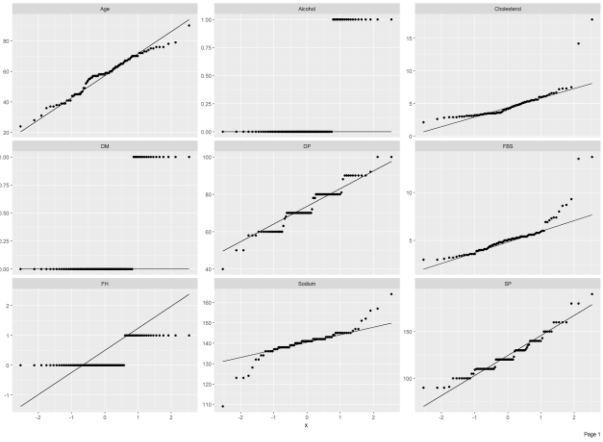


Figure 6: QQ Plot

4.5 Classification

Classification in Artificial intelligence is indeed a process that necessitates the application of machine learning techniques to learn how and where to assign a classifier to problem domain instances. In machine learning, there are many distinct types of categorization tasks to be encountered, as well as specialized modeling approaches to be employed for each. There is no effective model for mapping methods to problem categories; instead, practitioners are advised to conduct controlled tests to determine which algorithm configuration performs best for a given classification task and determining the accuracy of a classification algorithm requires Classification metric for evaluating a model's performance based on projected class labels. Nevertheless, this study presents in table 3 the risk factors for hypertension that can be categorized as modifiable and non-modifiable which forms the variables for the Machine Learning Classification.

Table 3: Hypertension Risk Metrics

Risk Factor	**Modifiability**	**Risk Description**
Age	Non-modifiable	HBP risk rises with age, notably after 65, due to arterial changes affecting elasticity
Systolic Pressure	Modifiable	High systolic blood pressure is a major risk for hypertension, indicating increased force on artery walls during heartbeats, raising cardiovascular and stroke risk
Diastolic Pressure	Modifiable	Increased diastolic blood pressure is a hypertension risk, indicating force on artery walls at rest. Raises heart disease and stroke risk
Family History	Non-modifiable	A familial hypertension history elevates high blood pressure risk, influenced by genetics, a non-modifiable risk factor.
Sodium Intake	Modifiable	Overconsumption of salt, a modifiable risk, raises blood pressure as the body retains water to dilute it.
Cholesterol	Modifiable	Elevated LDL cholesterol, a modifiable risk, contributes to arterial

		plaque buildup, raising the risk of hypertension and heart disease
Blood Sugar Level	Modifiable	Uncontrolled blood sugar levels from diabetes pose modifiable risks for hypertension, causing vascular and cardiac damage.
Alcohol Intake	Modifiable	Heavy alcohol consumption is a modifiable risk for hypertension, impacting blood pressure and medication effectiveness
Diabetic	Modifiable	Diabetes, a modifiable risk for hypertension, damages blood vessels and the heart, elevating the risk of complications.
Body Mass Index	Modifiable	Excess weight, a modifiable risk for hypertension, strains the heart and increases the risk of high blood pressure

4.5.1 SVM Classifier

Support vector machines are a collection of supervised learning methods for Classification, prediction, and anomaly analysis. The maximum margin classifier, or MMC, is the decision border established by SVMs. Making a straight line between two classes is how a simple linear SVM classifier works. Therefore, the SVM classifier result on this study is presented figure 7 which was trained using the training data.

```
Call:
svm(formula = Risk ~ ., data = train, kernel = "polynomial"

Parameters:
   SVM-Type:  C-classification
 SVM-Kernel:  polynomial
       cost:  1
     degree:  3
     coef.0:  0

Number of Support Vectors:  70

 ( 35 35 )


Number of Classes:  2

Levels:
 High Low
```

Figure 7: SVM model Summary

Again, this study further visualized how the SVM model important features during the training process the results are presented in figure 7a- 7c.

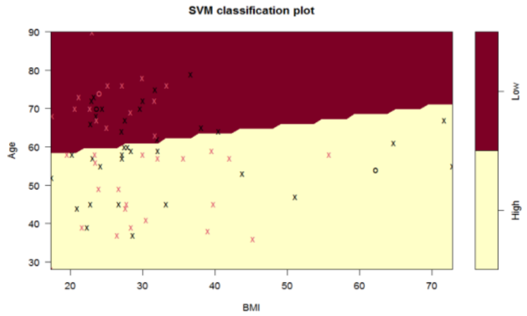


Figure 7a: Age Vs BMI

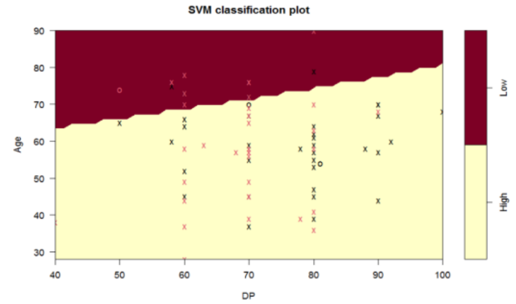


Figure 7b: Age Vs DP

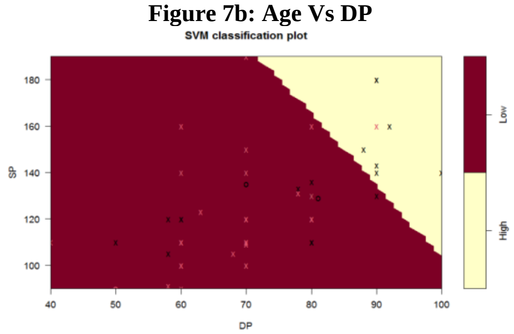


Figure 7c: SP Vs DP

4.5.2 Decision Tree Classifier

A decision tree classifier is an important classifier in the context of supervised learning, making its mark across statistical analysis, data mining, and machine learning. This algorithm shows its ability in tackling both classification and regression problems with. It uses the concept of a branching flowchart, where each internal node representing a distinctive feature, while the branches articulate the rules of decision-making, and the leaf nodes signify the ultimate outcomes derived from the algorithm's process. Therefore, the DT classifier result on this study is presented figure 8 which was trained using the training data.

```
> dec1=rpart(Risk~., data=train)
> dec1
n= 73

node), split, n, loss, yval, (yprob)
      * denotes terminal node

 1) root 73 36 High (0.5068493 0.4931507)
   2) SP>=126.5 31  9 High (0.7096774 0.2903226) *
   3) SP< 126.5 42 15 Low (0.3571429 0.6428571)
     6) Cholesterol< 3.45 12  5 High (0.5833333 0.4166667) *
     7) Cholesterol>=3.45 30  8 Low (0.2666667 0.7333333)
      14) SP>=109.5 23  8 Low (0.3478261 0.6521739)
        28) Cholesterol>=4.55 11  5 High (0.5454545 0.4545455) *
        29) Cholesterol< 4.55 12  2 Low (0.1666667 0.8333333) *
      15) SP< 109.5 7  0 Low (0.0000000 1.0000000) *
```

Figure 8: DT Trained model

Furthermore, figure 9 depicts the decision tree build from the training data this research.

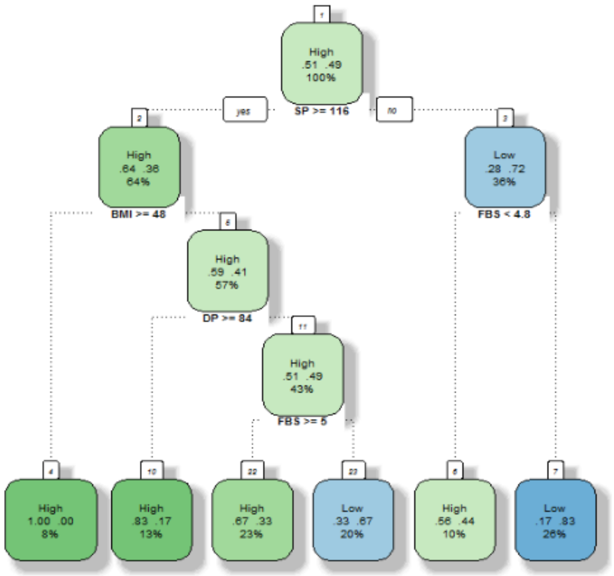
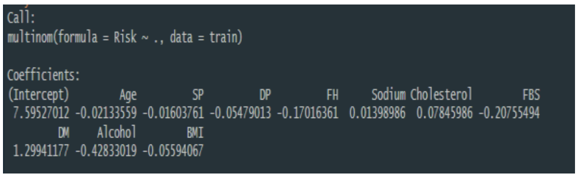


Figure 9: Decision Tree model

4.5.3 General Linear Model Classifier

Logistic regression also referred General linear model, a mainstay in supervised learning, tackles binary classification quandaries where outcomes are confined to two possibilities. Employing a sigmoid function, it gauges the likelihood of an input falling into a specific class. Through this mechanism, the linear regression's continuous output gracefully transitions into categorical values, all orchestrated by the sigmoid function's transformative power. The summary of the trained GLM is presented in figure 10



```
Call:
multinom(formula = Risk ~ ., data = train)

Coefficients:
(Intercept)         Age          SP          DP          FH  Sodium Cholesterol         FBS
 7.59527012 -0.02133559 -0.01603761 -0.05479013 -0.17016361  0.01398986  0.07845986 -0.20755494
         DM     Alcohol         BMI
 1.29941177 -0.42833019 -0.05594067
```

Figure 10: GLM Trained model

4.5.6 Performance Evaluation

Assessing the performance of machine learning models is vital to ascertain their precision, dependability, and utility. This critical process aids in selecting the most suitable model for a particular problem and ensuring its effectiveness, accuracy, and trustworthiness in real-world applications [19]. Particularly in sensitive domains like healthcare, errors or inaccuracies can have dire consequences, potentially resulting in fatalities. Utilizing the confusion matrix evaluation technique becomes imperative for model assessment [20]. This matrix, serving as a concise summary of a classification algorithm's performance, provides essential metrics such as accuracy, sensitivity, specificity, precision, and f-measure [21]. Accuracy measures the model's correctness by dividing the number of accurate predictions by the total predictions made. Sensitivity quantifies the proportion of actual positive cases correctly identified, while specificity measures the accuracy in identifying negative cases. Precision delves into the ratio of true positive cases among all positive predictions, while the f-measure represents the harmonic mean of precision and recall. In essence, the meticulous evaluation of machine learning models, facilitated by the confusion matrix method and its associated metrics, is indispensable for ensuring model efficacy and reliability. Hence using Confusion matrix, we present the performance of each model use in table 4 to 6. From table 4 to it is observed that SVM gave 0.9which is 90% accuracy, DT gave 0.83which is 83% accuracy, and GLM gave 0.64which is 64% accuracy respectively.

Table 4: SVM Performance

Support Vector Machine		
Metric	**Estimator**	**Estimate**
Accuracy	binary	0.909091
Sensitivity	binary	1
Specificity	binary	0.818182
Precision	binary	0.846154
F_measure	binary	0.916667

Table 4: DT Performance

Decision Tree		
Metric	**Estimator**	**Estimate**
Accuracy	binary	0.833333
Sensitivity	binary	1
Specificity	binary	0.666667
Precision	binary	0.75
F_measure	binary	0.857143

Table 6: DT Performance

General Linear Model		
Metric	**Estimator**	**Estimate**
Accuracy	binary	0.647059
Sensitivity	binary	0.888889
Specificity	binary	0.375
Precision	binary	0.615385
F_measure	binary	0.727273

Furthermore, a comparative performance analysis of the three ML classifiers use in this study is using the different performance metrics is presented in figure 11.

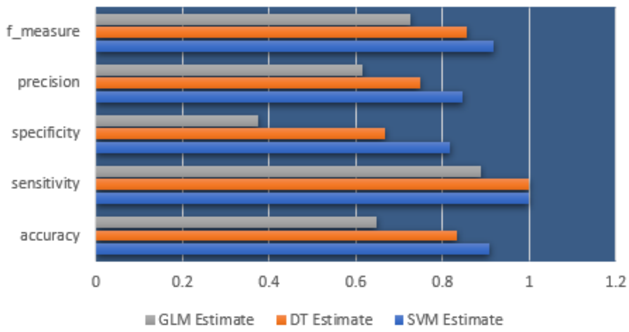


Figure 11: Performance Comparison

4.5.6 Model Deployment

Furthermore, using the best performed model that is the accuracy from the three models identify, SVM perform better. So SVM was deployed in a web-based interface were user at Realtime can supply hypertension parameter and used for the classification of hypertension risk conveniently. Hence figure 12 depict the interface deployed.

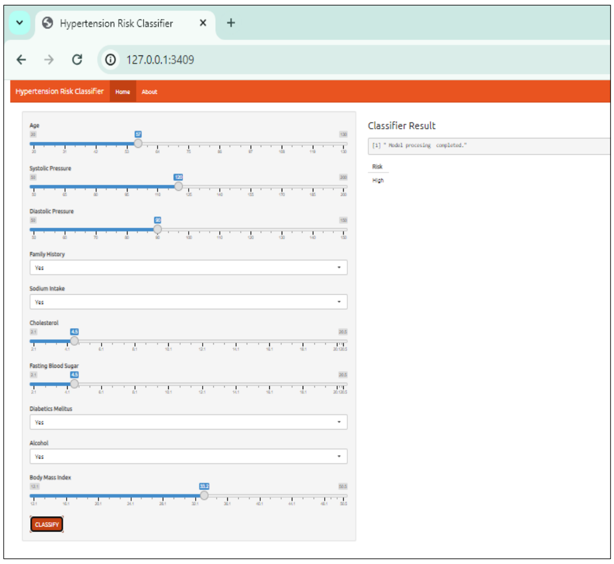


Figure 12: Deployed Interface

5. CONCLUSION

This study underscores the critical role of machine learning in enhancing the classification of hypertension risk, thereby aiding in early diagnosis and preventive interventions. Leveraging a Framework-Based Method, we curated a comprehensive dataset and applied various machine learning algorithms to classify hypertension risk effectively. Our findings reveal the superiority of Support Vector Machine (SVM) in accurately predicting hypertension risk, with an impressive accuracy rate of 90%. Additionally, Decision Tree and General Linear Model classifiers provided valuable insights, albeit with slightly lower accuracy rates. The deployment of SVM in a user-friendly web interface facilitates real-time risk assessment, empowering healthcare practitioners and individuals to make informed decisions. Future research endeavors should focus on refining machine learning models further, incorporating more extensive datasets from diverse populations to enhance generalizability. Additionally, exploring ensemble methods and deep learning techniques could improve classification accuracy and robustness.

6. ACKNOWLEDGMENT

All authors in this study discussed the results and contributed to the manuscript. Also, data used in this study is available on request.

7. REFERENCES

[1] K. T. Mills, A. Stefanescu, and J. He, "The global epidemiology of hypertension," Nature Reviews Nephrology, vol. 16, no. 4, pp. 223-237, Apr. 2020.

[2] Campbell, N. R. C., Paccot Burnens, M., Whelton, P. K., Angell, S. Y., Jaffe, M. G., Cohn, J., Espinosa Brito, A., Irazola, V., Brettler, J. W., Roccella, E. J., Maldonado Figueredo, J. I., Rosende, A., & Ordunez, P. (2022). 2021 World Health Organization guideline on pharmacological treatment of hypertension: Policy implications for the region of the Americas. Lancet regional health. Americas, 9, 100219. https://doi.org/10.1016/j.lana.2022.10021.

[3] Kjeldsen, S. E. (2018). Hypertension and cardiovascular risk: General aspects. *Pharmacological research*, 129, 95-99.

[4] Fuchs, F. D., & Whelton, P. K. (2020). High blood pressure and cardiovascular disease. *Hypertension*, 75(2), 285-292.

[5] Fuchs, F. D., & Whelton, P. K. (2020). High Blood Pressure and Cardiovascular Disease. *Hypertension (Dallas, Tex.: 1979), 75(2)*,285-292.

[6] Lackland DT (2010). Hypertension risk prediction: an important but complicated assessment. Vol 55(6):1304-5. Accessed from http://ahajournals.org by on October 26, 2023.

[7] Ramezankhani, A., Kabir, A., Pournik, O., Azizi, F., & Hadaegh, F. (2016). Classification-based data mining for identification of risk patterns associated with hypertension in Middle Eastern population: A 12-year longitudinal study. *Medicine*, 95(35), e4143.

[8] Li Y, Zhang G, Wang N. (2017). Systematic Characterization and Prediction of Human Hypertension Genes.

[9] Ye C, Fu T, Shiying, Zhang Y, Wang O, Jin Bo, Xia M, Modi Liu, Zhou X, Wu Q, Guo Y, Zhu C, Li Y, Culver D, Alfreds S, Stearns F, Sylvester K, Widen E, McElhinney D, Ling X (2018). Prediction of Incident Hypertension Within the Next Year: Prospective Study Using Statewide Electronic Health Records and Machine Learning. JOURNAL OF MEDICAL INTERNET RESEARCH. J Med Internet Res 2018 | vol. 20 | iss. 1 | e22 | p. 1.

[10] Duan T, Rajpurkar P, Laird D, Ng AY, Basu S. (2019). Clinical value of predicting individual treatment effects for intensive blood pressure therapy.Circ Cardiovasc Qual Outcomes. 2019; 12:e005010.

[11] Koshimizu, H., Kojima, R., Kario, K., & Okuno, Y. (2020). Prediction of blood pressure variability using deep neural networks. *International Journal of Medical Informatics*, 136, 104067.

[12] Rajput J.S., Sharma M., Tan R.S Rajendra A. (2020). Automated detection of severity of hypertension ECG signals using an optimal bi-orthogonal wavelet filter bank, Computers in Biology and Medicine.

[13] AlKaabi LA, Ahmed LS, Al Attiyah MF, Abdel-Rahman ME (2020) Predicting hypertension using machine learning: Findings from Qatar Biobank Study. PLoS ONE 15(10): e0240370.

[14] Essien V., Umoren I., Umoh I. A Bio-Informatics System for Intelligent Classification of Severity Index of Hypertension, International Journal of Innovative Research in Sciences and Engineering Studies (IJIRSES), 2021; ISSN: 2583-1658 | Volume: 1 Issue: 2, pp.1-8.

[15] Umoren, I., Abe, O., Ansa, G., Inyang, S., & Umoh, I. (2023). A New Index for Intelligent Classification of Early Syndromic of Cardiovascular (CVD) Diseases Based on Electrocardiogram (ECG). European Journal of Computer Science and Information Technology, 11(4), 1-21.

[16] Inyang, S., & Umoren, I. (2023). From Text to Insights: NLP-Driven Classification of Infectious Diseases Based on Ecological Risk Factors. *Journal of Innovation Information Technology and Application* (JINITA), 5(2), 154-165.

[17] Umoren, I. J., & Inyang, S. J. (2021). Methodical Performance Modelling of Mobile Broadband Networks

with Soft Computing Model. *International Journal of Computer Applications*, 174(25), 7-21.

[18] S. Inyang and I. Umoren, "Semantic-Based Natural Language Processing for Classification of Infectious Diseases Based on Ecological Factors," *International Journal of Innovative Research in Sciences and Engineering Studies (IJIRSES)*, vol. 3, no. 7, pp. 11- 21, 2023.

[19] Edet, A. E., & Ansa, G. O. (2023). Machine learning enabled system for intelligent classification of host-based intrusion severity. *Global Journal of Engineering and Technology Advances*, 16(03), 041-050.

[20] Ekong, A., Silas, A., & Inyang, S. (2022). A Machine Learning Approach for Prediction of Students' Admissibility for Post-Secondary Education using Artificial Neural Network. Int. J. Comput. Appl, 184, 44-49.

[21] Ekong, A., Udo, E., Ekong, O., & Inyang, S. (2023). Machine Learning based Model for the Prediction of Fasting Blood Sugar Level towards Cardiovascular Disease Control for the Enhancement of Public Health. International Journal of Computer Applications, 975, 8887.