International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868
Foundation of Computer Science FCS, Delaware, USA
Volume 13– No. 1, November 2025 – www.ijais.org

# A Machine Learning Approach to Predicting High Blood Pressure using Predictive Modeling on Local and Global Datasets to Enhance Patient Safety

Tolulope Esther Alabede
Department of ICT
Federal Medical Centre
Yenagoa PMb 502
Bayelsa, Nigeria

Japheth Richard Bunakiye
Department of Computer Science,
Niger Delta University,
Wilberforce Island,
Bayelsa, Nigeria

Michael Doorumun Ishima
Department of ICT,
Federal Medical Centre,
PMB 502 Yenagoa,
Bayelsa, Nigeria

Alimot Olaide Abdulazeez
ICT Department,
University College
Hospital, Ibadan, Nigeria

## ABSTRACT
Hypertension remains a major global public health burden, contributing to cardiovascular disease and premature deaths. Despite advances in medical care, delayed diagnosis particularly in developing countries like Nigeria continues to undermine effective prevention and intervention strategies. Traditional approaches rely on periodic measurements, which often fail to capture early risk indicators/patient-specific factors. With the growing availability of large-scale clinical data, machine learning provides an opportunity to enhance predictive modeling for early detection. This study proposed a machine learning framework for predicting hypertension using both local (426 patient records from Federal Medical Centre, Yenagoa) and global datasets (174,982 instances from Kaggle). The dataset was preprocessed using python libraries. Four ML algorithms: Logistic Regression, Random Forest, K-Nearest Neighbor, and XGBoost were trained separately on different feature dimensions with evaluation metrics including accuracy, sensitivity, specificity, F1-score, and AUC-ROC. Results indicated that RF achieved ~99.95% accuracy on the global dataset, while XGB on local data attained ~98.84% with superior sensitivity in distinguishing high-risk categories. A prototype web app built from the best-performing model was successfully tested, showing strong clinical potential. The study highlights that using local and global datasets improved generalization, while ensemble models enhanced predictive reliability for early detection to improve patient safety.

## General Terms
High Blood Pressure Prediction, Cardiovascular Risks

## Keywords
Hypertension, Blood Pressure, Machine Learning, Predictive Modeling, Systolic BP, Diastolic BP

## 1. INTRODUCTION
High blood pressure commonly known as hypertension is one of the non-communicable diseases characterized by an elevated blood pressure levels in the arteries. With every heartbeat, the heart pushes blood through the vessels, ensuring it reaches all parts of the body in a continuous cycle. The force of blood pressing against artery walls while it is pumped by the heart is what causes blood pressure. The illness, which is primarily defined by consistently elevated blood vessel pressure, makes it more difficult for the heart to pump blood. The vessels transport blood from the heart to every region of the body. With each heartbeat, the heart pumps blood into the vessels.

High Blood Pressure (hypertension), is a chronic medical condition in which the force of the blood against the artery walls is consistently too high. According to the World Health Organization [1], hypertension is diagnosed when an individual's blood pressure readings are consistently at or above 140/90 mmHg. However, some clinical guidelines, such as those from the American College of Cardiology (ACC) and the American Heart Association (AHA), define hypertension as a blood pressure reading of 130/80 mmHg or higher [2].

Hypertension is particularly challenging because it is an asymptomatic, silent killer and often remains hidden until caught during monitoring or evidenced in a hypertension-associated disease such as heart failure or stroke. When hypertension is not discovered or diagnosed, it leads to the increased risk of developing brain, kidney and cardiovascular diseases significantly. It also accounts for about half of all heart disease and stroke-related deaths worldwide [3].

According to WHO [1], it is a leading global health concern, affecting approximately 1.4 billion people worldwide. Major risk factors include a high-salt diet, obesity, physical inactivity, excessive alcohol use, and genetic predisposition. Effective management involves lifestyle modifications and, when necessary, antihypertensive medication.

Despite advancements in healthcare, Montagna et al. [4] stated that, current screening protocols exhibit high sensitivity but suffer from poor specificity, leading to unnecessary further assessments and increased healthcare costs. The World Hypertension League reports that in 2018, only about 59.5% of individuals with hypertension were aware of their condition, underscoring gaps in early detection and diagnosis.

In recent years, advances in data collection technologies, wearable health devices, and electronic health records (EHRs) have led to the accumulation of vast local and global health datasets. These datasets contain valuable insights that, when analyzed effectively, can support early detection and prediction of hypertension. Machine Learning (ML), a subfield of Artificial Intelligence (AI), has proven to be a powerful tool for modeling complex health data and uncovering hidden patterns that are not easily identifiable through traditional statistical methods.

Although several ML models have been developed for predicting high blood pressure, many are limited in scope, trained on small or non-diverse datasets, and not adequately validated across different populations. A critical need exists to create scalable and accurate ML models trained and tested on both local and global datasets to enhance prediction performance and ensure clinical relevance across different demographic and geographic contexts.

This research explores how ML techniques can be applied to predict hypertension by leveraging both local and global datasets. The goal is to improve patient safety by facilitating early risk detection, supporting proactive clinical decision-making, and promoting personalized preventive care.

## 2. RELATED WORKS

Several existing studies have explored the application of machine learning (ML) techniques in predicting hypertension and related cardiovascular conditions. The review focuses on methodologies, datasets, feature selection techniques, model performance, and their clinical implications. By examining these studies conducted across diverse populations and healthcare contexts, the review identified progress made, exposed prevailing limitations, and underscored gaps that this present research sought to address.

To begin with, Haruna [5] investigated hypertension prevalence in Jigawa State, Nigeria, by comparing the performance of five machine learning models: Random Forest (RF), Classification and Regression Trees (CART), Regression Tree (RT), Support Vector Machine (SVM), and Artificial Neural Networks (ANN). The ANN model outperformed the others with an AUC of 0.8694, revealing the importance of parental history of hypertension and diabetes as key predictors. This study underscores the significance of ML in localized public health applications, especially in identifying high-risk groups for targeted interventions.

Expanding on this, Mondal and Hazra [6] focused on early hypertension detection using datasets comprising physiological, demographic, and lifestyle data. They employed SVM, Decision Tree, and Logistic Regression, with genetic algorithms for feature selection. Their findings showed the Decision Tree model achieving the highest accuracy at 92.3%, highlighting its potential as a reliable diagnostic tool in clinical practice.

Similarly, Montagna et al. [4] leveraged World Hypertension Day datasets collected during health campaigns to train models on demographic, lifestyle, and physiological features. Using resampling techniques like SMOTE and undersampling, they evaluated supervised algorithms including RF, SVM, XGBoost, and KNN. Random Forest achieved the best balance of sensitivity (0.818) and specificity (0.629), outperforming traditional screening protocols, and emphasizing the role of data quality and class balance in enhancing predictive performance.

In another significant contribution, Obafemi [7] utilized Kaggle datasets and ensemble techniques including CatBoost, LightGBM, and Random Forest. These models demonstrated high performance, particularly CatBoost and LightGBM, which, after parameter tuning, achieved accuracy levels above 91% and RMSE scores as low as 0.87773. The study reinforces the strength of boosting algorithms in hypertension prediction.

Likewise, Effati et al. [8] employed ML models such as RF, XGBoost, SVM, and Logistic Regression using occupational health datasets. Feature selection using the k-best approach identified key predictors. The ensemble models, especially Random Forest, achieved accuracy rates between 97% and 99%, supporting the robustness of ensemble learning in high-risk population screening.

From a regional context, Kurniawan et al. [9] examined hypertension prediction among Indonesian adults using decision trees, random forest, gradient boosting, and logistic regression. Logistic regression outperformed others, yielding an AUC of 0.829, accuracy of 89.6%, and F1-score of 0.877. Their findings confirm the strong correlation between predictors and hypertension, validating logistic regression's efficacy for clinical applications.

In a broader cardiovascular context, Garg et al. [10] applied supervised learning models—Random Forest and K-Nearest Neighbor (KNN)—for heart disease prediction using features like age and cholesterol. KNN achieved a higher accuracy (86.89%) compared to Random Forest (81.97%), suggesting that KNN may be more adaptable for certain cardiovascular predictions.

Focusing on sub-Saharan Africa, Islam et al. [11] utilized data from 612 Ethiopian respondents and implemented the Boruta feature selection method with models including LR, ANN, RF, and XGBoost. The XGBoost model achieved the best results, with 88.81% accuracy and an AUC of 0.894, demonstrating ML's viability in low-resource settings for risk stratification.

Furthermore, Jeong et al. [12] developed predictive models using Korea's National Health Insurance data, analyzing hypertension incidence based on health check frequency. Their XGBoost model achieved an accuracy of 0.828 and an F1-score of 0.800, proving more effective than traditional logistic regression, and emphasizing the role of routine health screenings in disease forecasting.

Turning to model comparisons, Sandhiya et al. [13] evaluated ML models like LightGBM, CatBoost, XGBoost, and RF using features like cholesterol, heart history, and alcohol consumption. LightGBM emerged best with 92% accuracy, indicating that gradient boosting models are well-suited for complex clinical datasets due to their interpretability and performance.

In their attempt to build hybrid ML models, Das et al. [14] used Random Forest, Gaussian Naive Bayes, and SVC. Their Random Forest model yielded an 88% accuracy, with findings confirming that ensemble techniques offer robustness against noisy data and outperform single classifiers. For personalized hypertension risk prediction, Du et al. [15] developed a web-based system using SHAP explanations and ML models. LightGBM led with 70.57% accuracy. SHAP visualizations provided insights into risk factors, demonstrating the potential for personalized interventions through user-facing tools.

Cross-national validation was explored by Hwang et al. [16] who used ensemble models on South Korean and Japanese cohorts. Their Adaptive Boosting and logistic regression ensemble achieved an AUC of 0.901, with strong generalizability confirmed on the external dataset. This validates the feasibility of applying ML across borders using well-selected features like BMI and fasting glucose. In a maternal health application, Wanriko et al. [17] addressed pregnancy-induced hypertension in Kenya. With SMOTE

applied for class balancing, the Random Forest model achieved 89.62% accuracy, proving effective in managing maternal hypertension risks through early prediction.

Chang et al. [18] tackled hypertension complications by adopting a two-step feature elimination and classification process using models like SVM, C4.5, and XGBoost. XGBoost achieved 94.36% accuracy and an AUC of 0.927, proving its superior ability to reduce feature space while maintaining high performance, thus aiding clinical decisions for managing serious outcomes.

## 3. MATERIALS AND METHODS

The architecture of the proposed HBP prediction system is structured as a modular pipeline comprising five major stages as shown in Figure 1.

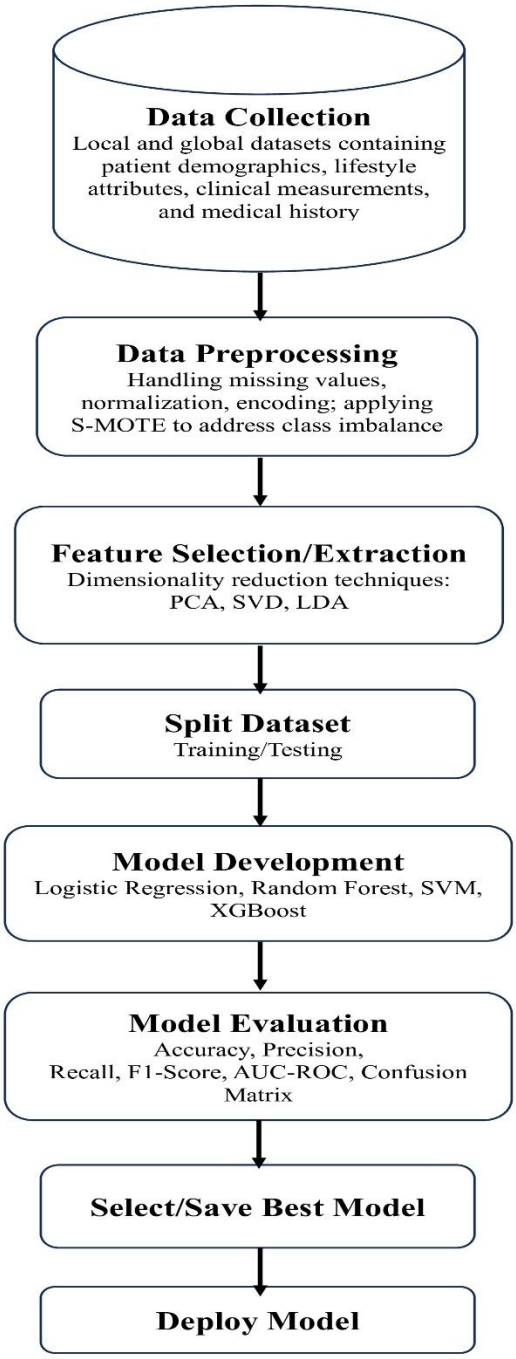


**Figure 1: Architecture of Proposed Model**

### 3.1 Data Collection

The success of any machine learning implementation relies heavily on the availability of a dataset [19]. Algorithms, including those used in this study, require data to generate accurate predictions. Consequently, data collection represents the critical first step in the machine learning process. The first

objective this study achieved was to collect a dataset of risks and indicators of hypertension. This study utilized two different datasets, a global dataset collected from a publicly available dataset repository, www.kaggle.com while the second dataset was collected locally containing vital signs data of patients from the clinical historical records of patients across outpatient clinics in Federal Medical Centre, Yenagoa, Bayelsa State after obtaining due ethical approval from the hospital's ethics committee with reference FMCY/REC/ECC/2025/AUGUST/898.

## 3.2 Data Preprocessing

Prior to model training, the datasets required thorough pre-processing, as input samples contained diverse features and often included missing or inconsistent values. Since many missing entries were recorded measurements from staff, data-cleaning techniques were employed to correct errors and inconsistencies, including duplicates, outliers, and absent values. This cleaning process utilized various methods such as imputation, removal, and data transformation to ensure the dataset's integrity and suitability for machine learning.

When initiating the creation of a machine learning model, data preprocessing serves as the foundational step. Real-world data often presents challenges such as incompleteness, inconsistency, inaccuracy (including errors or outliers), and missing attribute values or trends. This is where preprocessing plays a vital role. Leveraging Exploratory Data Analysis (EDA) and Python libraries, it cleans, formats, and organizes raw data, ensuring it is properly structured and ready for machine learning applications [20].

The target variable for the analysis was constructed using the blood pressure classification system recommended by the ACC/AHA, which considers both systolic and diastolic measurements. Other features derived from the datasets which were vital indicators for predicting hypertension included BMI and Pulse Pressure.

## 3.3 Feature Selection/Extraction

In the development of a robust machine learning model for high blood pressure prediction, feature extraction plays a vital role in transforming raw input data into a more meaningful and compact form that enhances model performance. This process is essential for capturing the most informative characteristics of the data while minimizing redundancy and noise. Given the nature of clinical datasets, which often contain numerous correlated or less-informative variables, the application of dimensionality reduction techniques becomes especially critical.

The objective of feature extraction is to simplify the dataset by generating a new set of features from the original ones. This reduced set is designed to preserve most of the essential information from the initial data. Through the combination and transformation of the original attributes, a more compact representation of key features is achieved, effectively capturing the underlying patterns and information [21].

Dimensionality reduction is a subset of feature extraction that involves reducing the number of input variables or features while preserving as much relevant information as possible. This not only improves computational efficiency but also helps mitigate the risks of overfitting and multicollinearity in the predictive modeling process. In this research, three dimensionality reduction techniques were employed based on their theoretical strengths and effectiveness observed in related works: Principal Component Analysis (PCA) and Singular Value Decomposition (SVD).

These feature extraction techniques will be applied after the data is preprocessed and normalized; ensuring that the reduced feature sets retained the most discriminative and clinically relevant information. The extracted features were then used as input into various machine learning algorithms including Logistic Regression, Random Forest, Support Vector Machine, and XGBoost.

By reducing the feature space effectively, the models were able to learn more generalizable patterns, improve prediction accuracy, and operate more efficiently, especially on the combined local and global datasets. The use of dimensionality reduction thus aligns with the research objective of identifying and isolating key predictive indicators of high blood pressure for enhanced patient safety.

## 3.4 Model Building

Model building is a critical phase in the machine learning pipeline, involving the selection, training, and tuning of appropriate algorithms to develop a predictive system capable of accurately identifying high blood pressure cases based on the features extracted from both local and global datasets. The goal is to construct models that generalize well on unseen data while maximizing key performance metrics such as accuracy, sensitivity, specificity etc.

In this study, four widely-used supervised machine learning algorithms were adopted: Logistic Regression (LR), Random Forest (RF), K-Nearest Neighbors and Extreme Gradient Boosting (XGBoost). These algorithms were selected based on their proven performance in medical prediction tasks and their unique strengths in handling classification problems.

The training process involved splitting the datasets into training and testing subsets using stratified sampling to maintain class distribution. Hyperparameter tuning was performed ensuring that each model was optimized for best performance. The reduced and transformed feature sets obtained through PCA and truncSVD were used as input for model training, enabling efficient learning from the most informative features.

### *3.4.1 Logistic Regression (LR)*

Logistic regression is a supervised machine learning algorithm in data science. It is a type of classification algorithm that predicts a discrete or categorical outcome. Logistic regression, like linear regression, is a type of linear model that examines the relationship between predictor variables (independent variables) and an output variable which is the response, target or dependent variable [16].

Logistic regression according to Jurafsky & Martin [22], can be used to classify an observation into one of two classes (like 'positive sentiment' and 'negative sentiment'), or into one of many classes.

According to Banoula [23], the term "logistic regression" comes from the idea of the logistic function it utilizes. The sigmoid function is another name for the logistic function. This logistic function has a value between 0 and 1. The simplicity of logistic regression is one of its primary benefits. In addition to making predictions, logistic regression in machine learning assists in determining which variables are most crucial to these forecasts. Because of this, logistic regression is a useful technique for resolving categorization issues and offering insightful information about the data. It is often used in

machine-learning projects due to its interpretability and simplicity of use. Binary logistic regression, multinomial logistic regression, and ordinal logistic regression are the three primary forms of logistic regression [23].

### *3.4.2 Random Forest (RF)*

Random Forest is an ensemble learning technique widely used for classification, regression, and related tasks. The algorithm constructs multiple decision trees using random subsets of both the training data and features, and combines their outputs to generate a final prediction. Compared to individual decision trees, Random Forest offers enhanced accuracy and greater resilience to overfitting. It is also capable of managing high-dimensional datasets and handling noisy or incomplete data. However, these benefits come at the cost of increased computational complexity and longer processing times relative to simpler models [24].

### *3.4.3 Extreme Gradient Boosting (XGBoost)*

XGBoost (eXtreme Gradient Boosting) is a distributed, open-source machine learning library that uses gradient boosted decision trees, a supervised learning boosting algorithm that makes use of gradient descent. It is known for its speed, efficiency and ability to scale well with large datasets [25].

XGBoost, or eXtreme Gradient Boosting, according to Tyagi [26], has become a popular choice for supervised learning tasks, including both regression and classification. The algorithm constructs a predictive model by aggregating the outputs of multiple base models, commonly decision trees, in an iterative fashion. Each successive model, or weak learner, is trained to correct the errors of the preceding ensemble members. During training, XGBoost employs gradient descent optimization to minimize a specified loss function, enhancing overall model accuracy and performance.

Tyagi [26] added that key features of XGBoost Algorithm include its ability to handle complex relationships in data, regularization techniques to prevent overfitting and incorporation of parallel processing for efficient computation.

### *3.4.4 K-Nearest Neighbors (KNN)*

K-Nearest Neighbors (KNN), introduced by Cover and Hart in 1967, is a machine learning algorithm employed for both classification and regression tasks. It generates predictions based on the similarity between observations, considering the k closest neighbors to a given data point. In classification tasks, the observation is assigned to the class most frequently represented among its nearest neighbors [27].

KNN as described by [19], is a simple instance-based technique that classifies unlabeled data points by referencing the nearest instances in the training set. Unlike decision tree algorithms, which build explicit tree structures, instance-based methods like KNN rely directly on the training data for prediction. Some researchers suggest that all learning algorithms are inherently instance-based, as they utilize training data to develop models. In KNN, classification involves calculating the distance between an unlabeled point and its neighbors using a defined distance metric, then assigning the point to the class most frequently represented among these neighbors. For example, in website classification, KNN measures the distance between the features of an unlabeled website and those of labeled instances, assigning the site to the most common class among its nearest neighbors. Known as a “lazy learning” algorithm, KNN is suitable for both regression and classification tasks.

## 3.5 Model Evaluation

Model evaluation is a critical process for determining the performance and generalizability of machine learning models. Evaluation metrics, as described by Srivastava [28], are quantitative measures that assess a model’s effectiveness. They offer insights into model performance and facilitate comparisons between different models or algorithms, guiding the selection of the most appropriate approach for a given task.

Srivastava further explains that when assessing a machine learning model, it is important to examine its prediction accuracy, ability to generalize, and overall performance. The selection of evaluation metrics should be guided by the particular problem area, data characteristics, and intended goals.

The research addressed a multi-classification problem and hence different classification metrics were applied to determine models’ performance. In the context of high blood pressure prediction, it is critical to use a comprehensive set of evaluation metrics to determine how well the model distinguishes between the different class labels, especially given the class imbalance often present in medical datasets.

Some of the most frequently used evaluation metrics for classification tasks that were used in this research include accuracy, precision, recall, confusion matrix, AUC-ROC and F1-score. It is considered good practice to evaluate a model with several different metrics, as this approach offers a more complete understanding of how well the model fits the specific problem it aims to address.

### *3.5.1 Confusion Matrix*

For machine learning classification problems where the output can be two or more classes, a confusion matrix is a performance measurement that is very helpful for measuring precision and recall, specificity, accuracy, and most importantly, AUC-ROC curves. A confusion matrix is a N X N matrix, where N is the number of predicted classes. For the problem at hand, we have N=2, so we get a 2 X 2 matrix Srivastava [28].

In a confusion matrix, the rows typically stand for the actual classes of the data points, while the columns correspond to the predicted classes, or vice versa. It can be used for both binary and multi classification problems. Table 1 shows the components and following are the terms that describes the confusion matrix

Important Terms in a Confusion Matrix:

1. **TP:** The benign URLs are correctly identified as benign.
2. **FP:** The benign URLs are incorrectly classified as malicious.
3. **TN:** The malicious URLs are accurately recognized as malicious.
4. **FN:** The malicious URLs are mistakenly identified as benign.

The confusion matrix provides a foundation for calculating various metrics to assess the classification model’s performance, with the choice of metrics depending on the specific requirements of the application domain.

**Table 1: Confusion matrix**

| | | Predicted Value | |
|---|---|---|---|
| | | *Positive* | *Negative* |
| Actual Value | *positive* | TP<br>True positive | FN<br>False negative |
| | *negative* | FP<br>False positive | TN<br>True negative |

### *3.5.2 Accuracy*

Most common evaluation metric that is used is accuracy. It measures how many observations both positive and negative, were correctly classified. The classification accuracy is the ratio of the number of correct predictions to the total number of input samples [19].

$$\boldsymbol{Accuracy} = \frac{No.of\ correct\ predictions}{Total\ No.of\ predictions\ made} \times 100 \quad (1)$$

But we can write this in terms of true positive, false positive etc. like this:

$$\boldsymbol{Acc} = \frac{TP + TN}{TP + FP + TN + FN} \quad (2)$$

### *3.5.3 Precision*

Precision (pr) evaluates a model's capacity to accurately classify positive instances. It calculates the number of the instances predicted as positive are actually positive Schlosser et al. [29]. In other words, it tells us how precise or confident the model is when labeling an instance as positive. High precision means fewer false positives, indicating the model is good at correctly identifying true positives and is computed as:

$$pr = \frac{TP}{TP + FP} \quad (3)$$

### *3.5.4 Recall*

Recall is the percentage of correctly predicted instances among all actual positive classes. High recall means the model successfully identified most of the true positives, even if it includes some false positives Swaminathan et al. [30] and it can be expressed using:

$$r = \frac{TP}{TP + FN} \quad (4)$$

### *3.5.5 AUC-ROC*

The Area Under the Receiver Operating Characteristic Curve (AUC–ROC) is a widely used metric for evaluating classification models across different threshold settings. The ROC curve is a probability plot that illustrates the trade-off between the true positive rate (TPR) on the y-axis and the false positive rate (FPR) on the x-axis. The AUC quantifies the model's ability to distinguish between classes, with higher values indicating better performance. In a medical context, a higher AUC reflects a model's greater capacity to correctly differentiate between patients with a disease and those without [31].

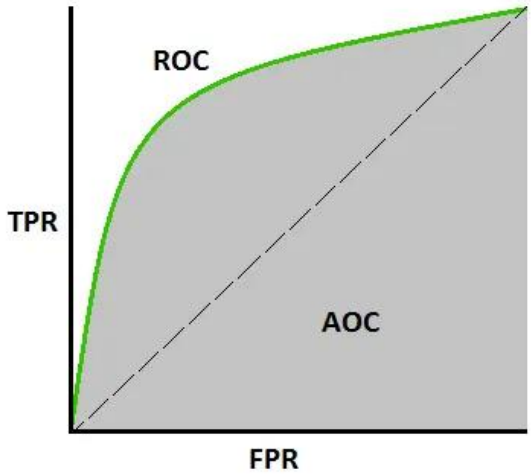


**Figure 2: AUC - ROC Curve [31]**

## 4. RESULTS AND DISCUSSION

### 4.1 Experimentation

The experimentation process to build this predictive model was conducted separately on both the Global hypertension dataset and the Local hypertension dataset following a systematic ML pipeline. In the first step, data preprocessing was carried out using EDA and missing values were handled. In order to ensure data remained uniform across the datasets, StandardScaler was used to standardize numerical features while categorical variables were encoded using One-Hot Encoding. Additionally, new features BMI and Pulse Pressure were derived from the clinical readings of the local dataset and were incorporated before the target variable to enhance predictive power. The target labels were directly assigned or transformed according to the four WHO blood pressure categories. Less informative attributes such as educational level, country, and employment status were excluded from the global datasets to avoid noise in the models.

After preprocessing, the dataset was split into training and testing sets using the 80:20 ratios with stratification to preserve the class distribution across the four hypertension categories. To handle class imbalance, the SMOTE technique was applied to the training set, generating synthetic minority class examples. Best results were obtained from experimenting separately on the original feature dimension, PCA and TruncSVD feature dimensions to Logistic Regression (LR), Random Forest (RF), Extreme Gradient Boosting (XGBoost) and KNN machine learning algorithms. Logistic Regression was used as a baseline linear classifier, KNN as a distance-based learning method, while Random Forest and XGBoost represented ensemble learning methods capable of capturing complex, nonlinear interactions. Each model was tuned using RandomizedSearchCV, where hyperparameter search spaces were defined for maximum depth, number of estimators, learning rates, and regularization strengths. To optimize computational efficiency, the randomized search was limited to 20 iterations with 3-fold cross-validation. Model evaluation was performed on the held-out test set using multiple metrics: Accuracy, Precision, Recall, F1-score, and Area Under the ROC Curve (AUC). Confusion matrices were plotted for each model to visualize classification performance across the four WHO categories of hypertension. Additionally, ROC-AUC curves were averaged per model to evaluate multi-class separability. To compare performance across models, bar

charts were generated for all metrics, and the best model was identified during tuning. The experimentation was implemented using Python (scikit-learn, XGBoost, imbalanced-learn, and Seaborn/Matplotlib libraries) within a Jupyter Notebook environment.

## 4.2 Results of ML Algorithms on Global Datasets

Three different feature sets (Original, PCA, SVD) were experimented on the global dataset and Table 2 summarizes the predictive performance of the four machine learning models which include Logistic Regression, Random Forest, XGBoost, and KNN. The results show that Random Forest and XGBoost on original features performed excellently having Accuracy, Precision, Recall, F1-Score > 99.90%, ROC-AUC = 100%. In contrast, Logistic Regression and KNN demonstrated more moderate performance. Dimensionality reduction with PCA and SVD generally led to a decrease in performance of the models, indicating that the original feature set contained the most discriminative information.

**Table 2: Model Performance Summary on Global Datasets**

| Model | Feature Set | Accuracy | Precision | Recall | F1-Score | ROC-AUC |
|---|---|---|---|---|---|---|
| Logistic Regression | Original Features | 0.79944 | 0.863487 | 0.79944 | 0.823491 | 0.940466 |
| Random Forest | Original Features | 0.999486 | 0.999488 | 0.999486 | 0.999484 | 1.000000 |
| XGBoost | Original Features | 0.9994 | 0.999403 | 0.9994 | 0.999398 | 1.000000 |
| KNN | Original Features | 0.797726 | 0.865027 | 0.797726 | 0.822796 | 0.888273 |
| Logistic Regression | PCA Features | 0.79944 | 0.863487 | 0.79944 | 0.823491 | 0.940466 |
| Random Forest | PCA Features | 0.953196 | 0.957974 | 0.953196 | 0.954638 | 0.995531 |
| XGBoost | PCA Features | 0.972283 | 0.974456 | 0.972283 | 0.972909 | 0.998439 |
| KNN | PCA Features | 0.797726 | 0.865027 | 0.797726 | 0.822796 | 0.888273 |
| Logistic Regression | SVD Features | 0.757408 | 0.83659 | 0.757408 | 0.788792 | 0.912082 |
| Random Forest | SVD Features | 0.754665 | 0.828731 | 0.754665 | 0.783635 | 0.905154 |
| XGBoost | SVD Features | 0.73472 | 0.831637 | 0.73472 | 0.772529 | 0.905691 |
| KNN | SVD Features | 0.695802 | 0.805998 | 0.695802 | 0.739271 | 0.793254 |

### *4.2.1 Random Forest Confusion Matrix on Original Feature Dimension*

The confusion matrix for the Random Forest model on the original features shows an extremely high number of correct predictions along the diagonal. The number of misclassifications is nearly zero, visually confirming the model's almost 100% accuracy in classifying all four blood pressure categories on the global test set. This is visualized in the confusion matrix in Figure 3.

### *4.2.2 ROC-AUC on Random Forest RF*

The Receiver Operating Characteristic (ROC) curves for the Random Forest model, using a One-vs-Rest approach for multi-class classification, showed in Figure 4 that the curves for each class (Normal, Elevated, Stage 1, Stage 2) are tightly aligned to the top-left corner. This indicates excellent class-wise separability and is consistent with the perfect macro-average ROC-AUC score of 1.000 reported in the performance summary.

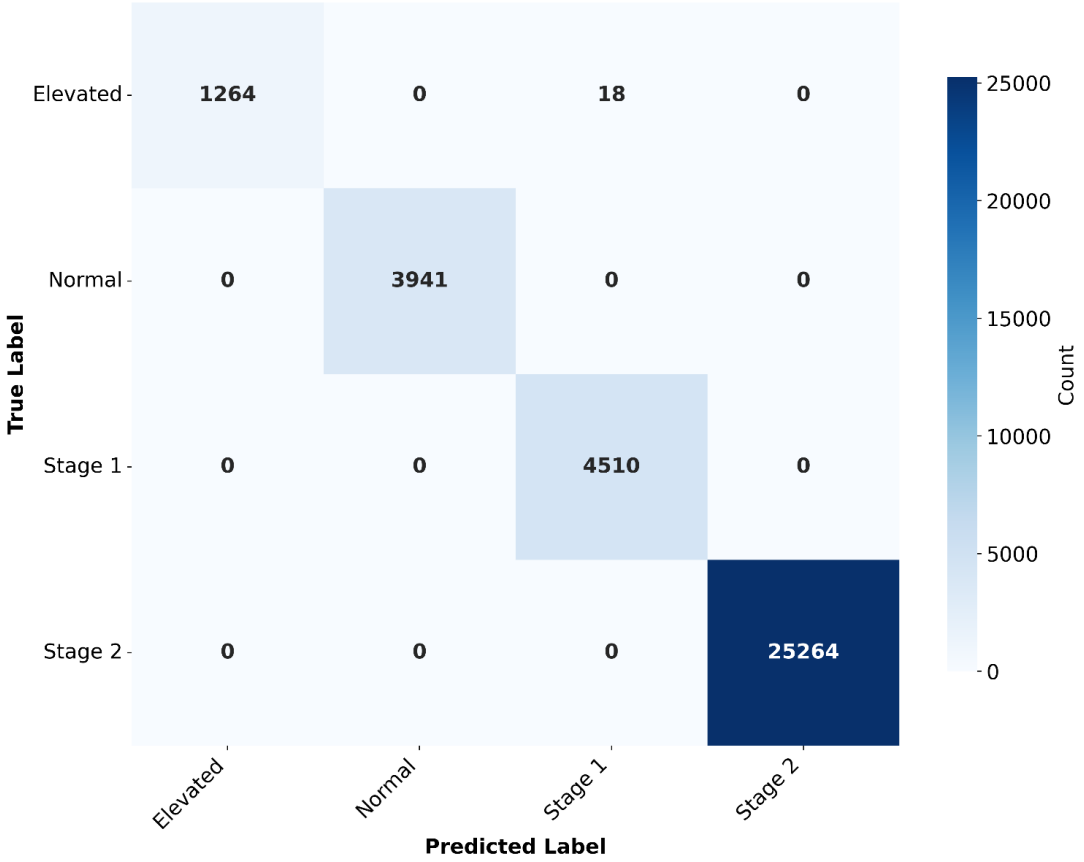


**Figure 3: Random Forest Confusion Matrix on Original Feature Dimensions**

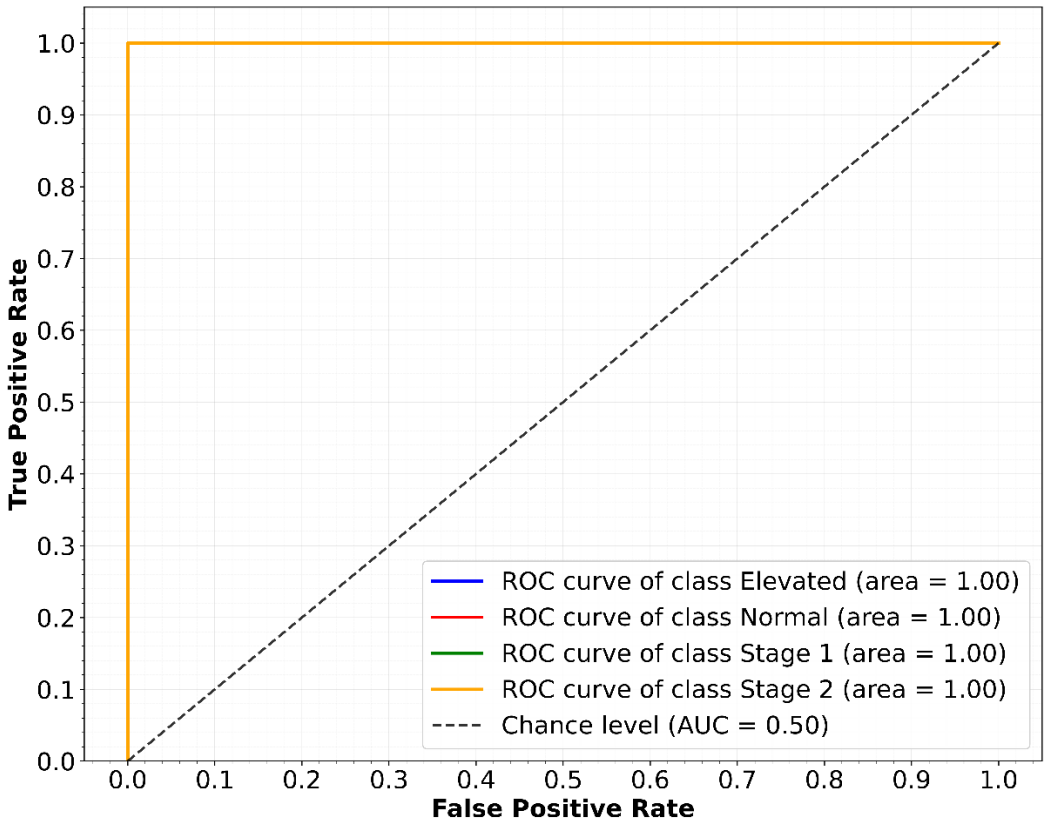


**Figure 4: ROC Curves – Randon Forest (Original Features)**

## 4.3 Results of ML Algorithm on Local Dataset

The same four models were trained and evaluated on the local dataset (FMC Yenagoa dataset) which is smaller in size than the global. As shown in Table 3 the Random Forest and XGBoost again outperformed other models, achieving identical and very high accuracy and F1-scores of 0.988. Logistic Regression performed well (0.930 accuracy), while KNN showed significantly lower performance of 0.756 accuracy. The ROC-AUC scores for the top models remained exceptionally high (0.99959 for RF, 1.000 for XGBoost), demonstrating strong predictive capability even with a limited sample size.

The bar chart in Figure 5 provides a visual comparison of the models' performance metrics on the local dataset. It clearly illustrates the superiority of Random Forest and XGBoost across all evaluated metrics (Accuracy, Precision, Recall, F1-Score, and AUC-ROC) compared to Logistic Regression and KNN.

**Table 3: Classification Summary of models on Local Datasets**

| Model | Accuracy | Precision | Recall | F1-Score | AUC-ROC |
|---|---|---|---|---|---|
| RandomForest | 0.988372 | 0.988735 | 0.988372 | 0.988355 | 0.99959 |
| XGBoost | 0.988372 | 0.988735 | 0.988372 | 0.988355 | 1.00000 |
| LogisticRegression | 0.930233 | 0.933852 | 0.930233 | 0.925404 | 0.98735 |
| KNN | 0.755814 | 0.805772 | 0.755814 | 0.756253 | 0.93057 |

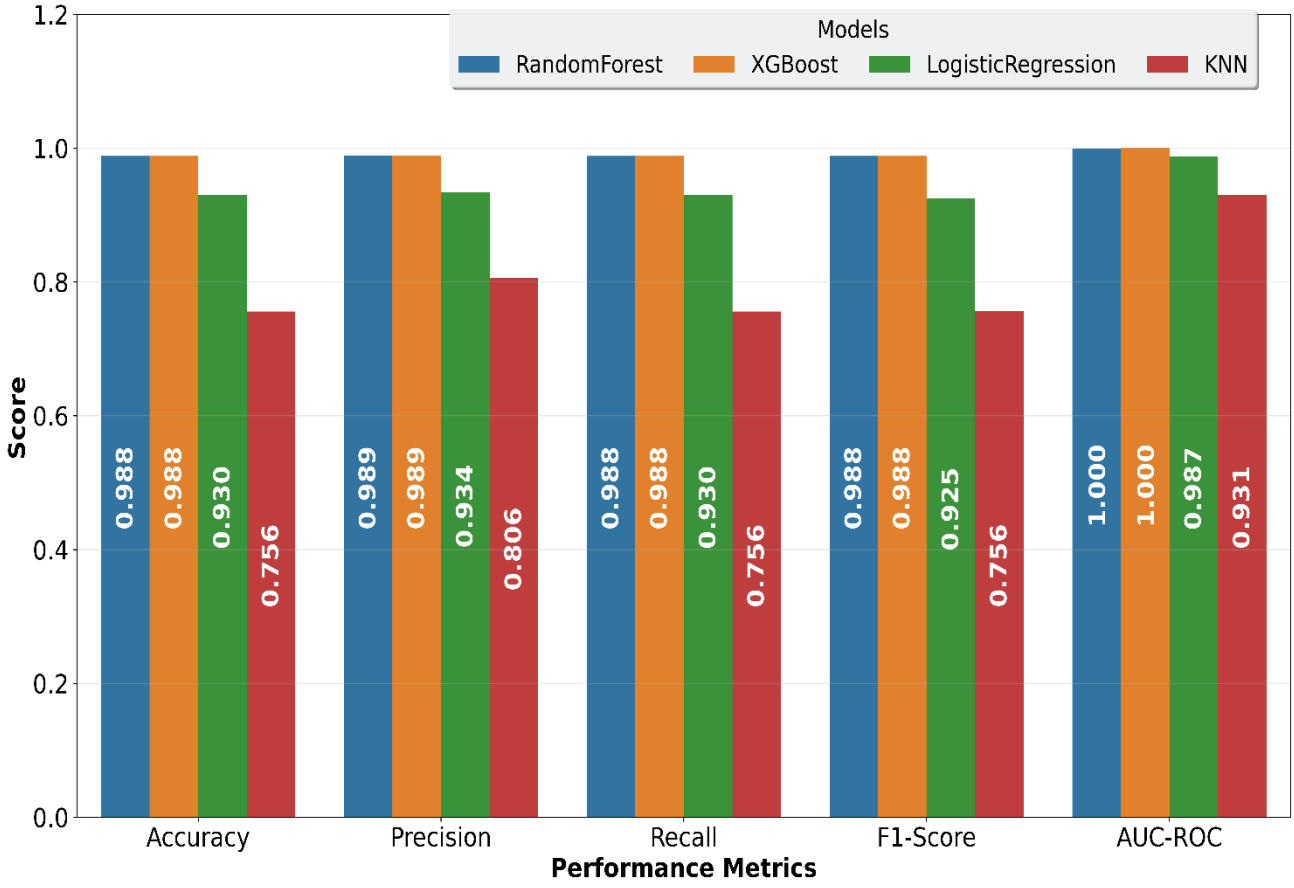


**Figure 5: Chart Showing Model Performance on Local Dataset**

### *4.3.1 XGBoost Confusion matrix results*

The confusion matrix for the XGBoost model on the local dataset as visually demonstrated in Figure 6 showing a strong diagonal, indicating most test instances were correctly classified. A small number of misclassifications occur between adjacent categories (e.g., Normal being predicted as Elevated), which is a common and clinically understandable error given the continuity of blood pressure values.

### *4.3.2 ROC-AUC Model Comparison*

The ROC curves for all four models on the local dataset visually confirm the results from the performance table in Figure 7. The ROC curves for Random Forest (RF) and

XGBoost (XGB) are closest to the top-left corner, indicating their superior performance. Logistic Regression (LR) follows, while KNN's curve is significantly lower, reflecting its lesser ability to distinguish between the classes.

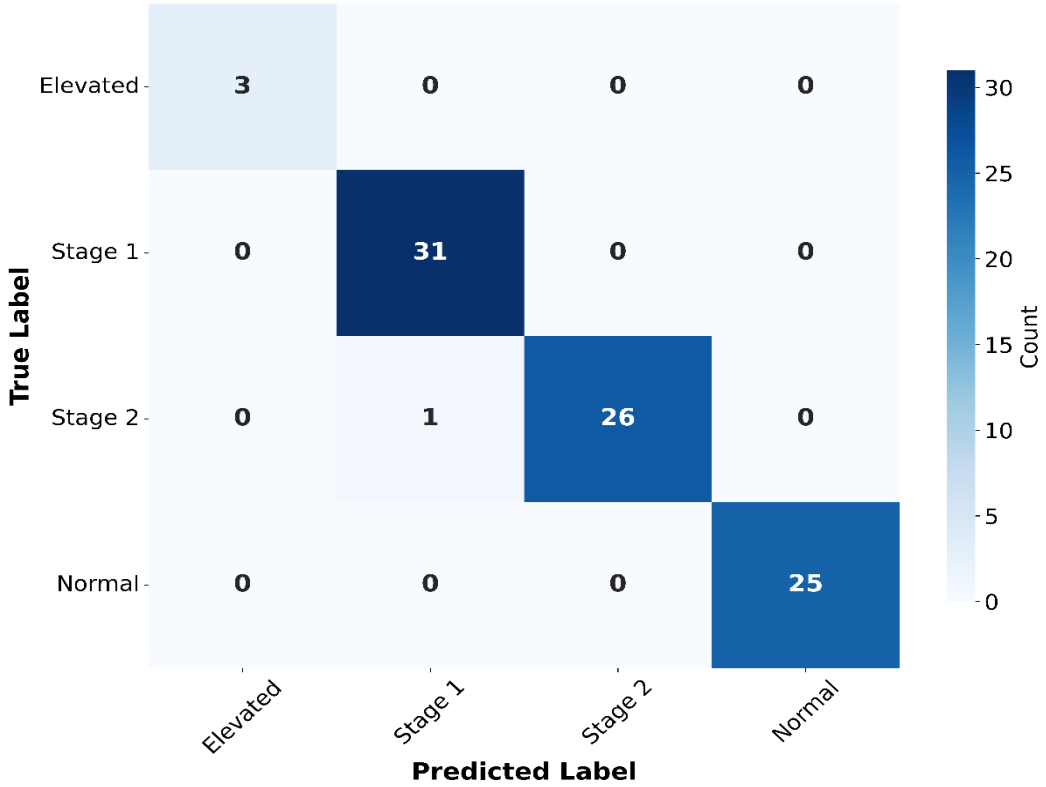


**Figure 6: XGBoost Confusion Matrix on Original Feature Dimension**

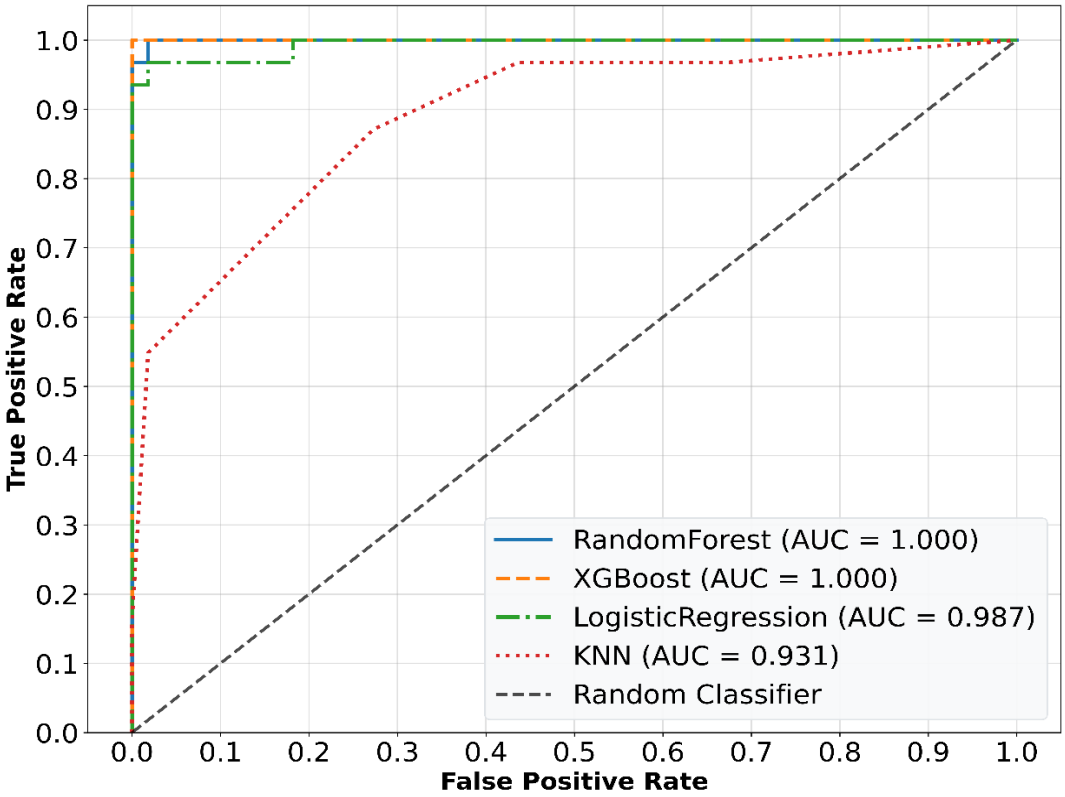


**Figure 7: ROC curves on RF, XGB, LR, KNN**

## 4.4 Comparing Improved vs Existing Models

The performance scores in this research are compared to five existing models in relevant and related studies. This evaluation aims to validate the effectiveness of the improved model in classifying and predicting different hypertension categories. State-of-the-art models were selected for comparison based on their relevance to this study and the improved model's ability to replicate or utilize their approaches, along with parameter optimizations such as cross-validation and the use of evaluation metrics. The improved prediction model demonstrated the highest performance among the five existing models. Table 4 presents a comparative summary of the performance of the improved model against the selected state-of-the-art models for hypertension prediction.

The chart in Figure 8 and Table 4 provides a compelling follow-up to the comparative analysis by highlighting the strong performance of the proposed prediction model. This reinforces the notion that the model's superior accuracy stems from its ability to overcome limitations found in previous models. Notably, the proposed approach places greater emphasis on meticulous feature selection, thorough data cleansing and preprocessing techniques which are critical steps that earlier studies may sometimes have neglected. This tailored approach ultimately contributes to the observed high accuracy and effectiveness relative to existing models.

**Table 4: Comparative summary of Improved vs Existing models**

| Author | Accuracy Results (%) |
|---|---|
| Mondal and Hazra (2024) | 92.3 |
| Montagna et al. (2022) | 74.7 |
| Obafemi (2022) | 91 |
| Kurniawan et al. (2023) | 89.6 |
| Du et al. (2023) | 70.57 |
| Proposed Model | 99.95 |

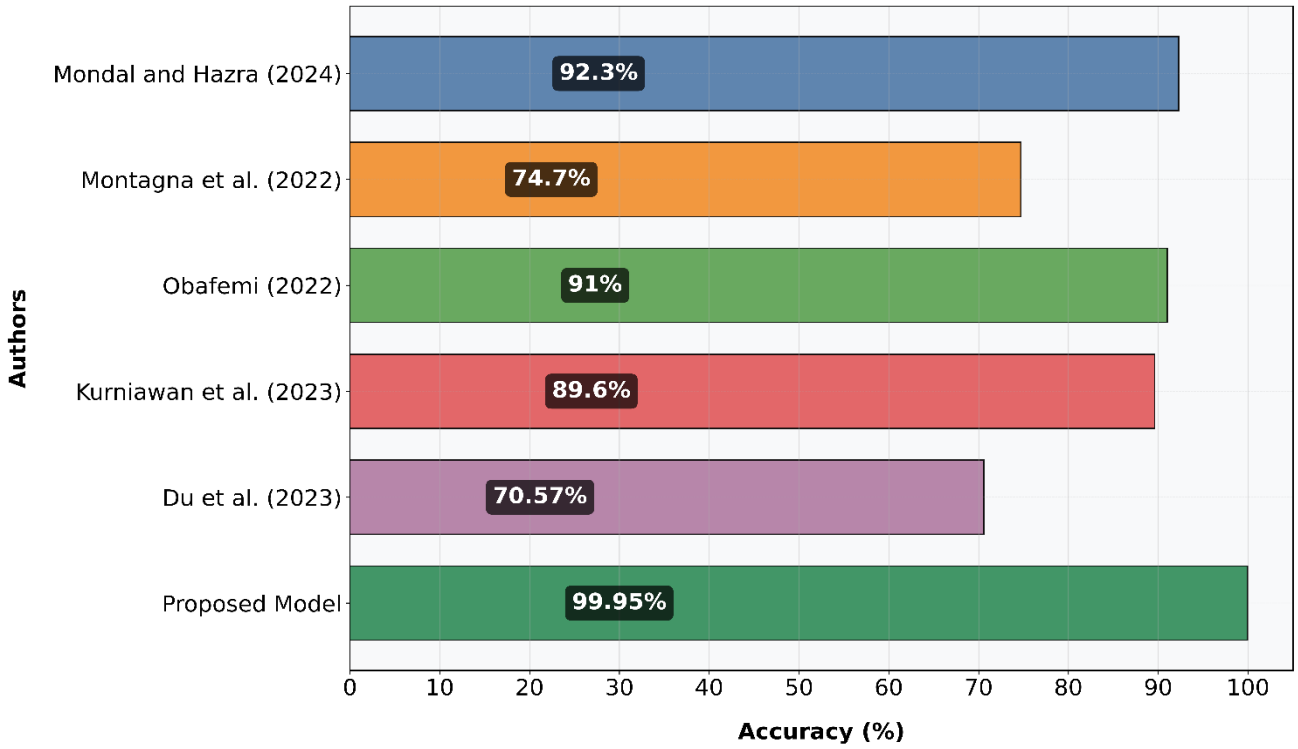


**Figure 8: Visualization of the Improved Model and the Existing Model(s)**

The results demonstrate that ensemble learning methods (Random Forest and XGBoost) are highly effective for hypertension prediction across both global and local contexts. Their ability to capture complex, non-linear relationships in the data makes them suitable for medical classification tasks where multiple interacting risk factors are present.

The near-perfect performance on the global dataset may be attributed to its large size and diversity, which allows models to learn generalized patterns effectively. However, the risk of overfitting must be considered, especially given the perfect scores. Cross-validation and hyperparameter tuning helped mitigate this, but further validation on external datasets is recommended.

On the local dataset, the high performance of Random Forest and XGBoost despite the small sample size highlights their robustness in data-scarce environments. This is particularly valuable in clinical settings in regions like Nigeria, where data collection is challenging but model accuracy is critical. The lower performance of KNN on both datasets underscores its limitations with high-dimensional data and class imbalance, making it less suitable for this type of predictive task without significant preprocessing.

## 4.5 Model Deployment

After training, evaluating, and tuning the models, the best-performing model was exported for real-world use. The trained model was serialized and saved using standard Python joblib libraries, allowing it to be reloaded without retraining. This ensured that the optimized parameters and learned patterns from the training phase were preserved for future predictions.

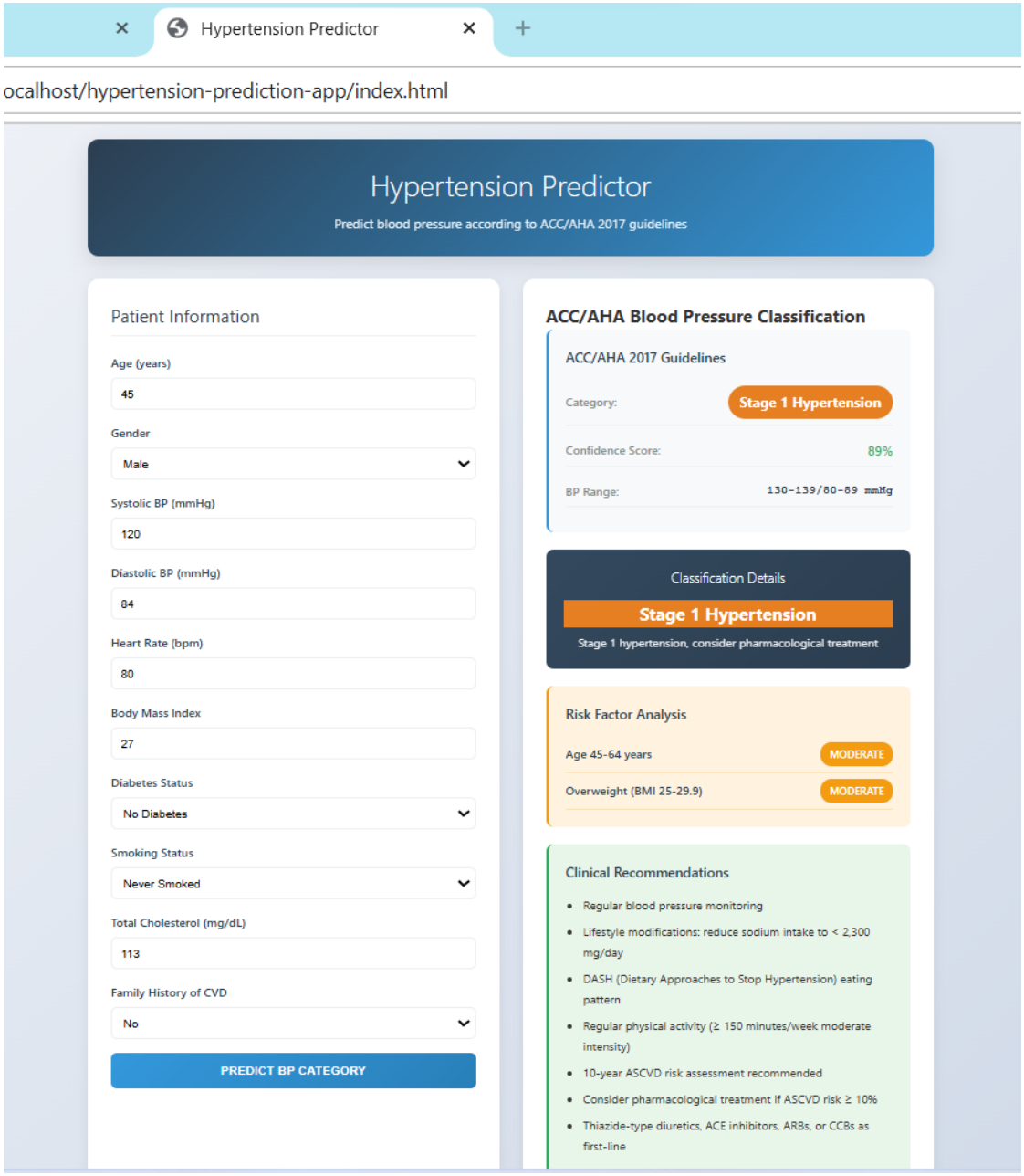


**Figure 9: Prediction results generated based on the input data**

To demonstrate its practical application, a simple web application prototype was developed, "Hypertension Predictor". The application provides an interactive interface where users can input health-related attributes such as age, gender, systolic and diastolic blood pressure, body mass index (BMI), family history, and other clinical variables as shown in Figure 9 Once submitted, the system processes the input through the saved model, predicts and outputs the corresponding hypertension category (Normal, Elevated, Stage 1 Hypertension, or Stage 2 Hypertension) as Figure 9 shows. Also included on the prediction output page are clinical recommendations for the various hypertension stages.

# 5. CONCLUSION

This application of both local and global datasets for predicting hypertension categories demonstrates the high potential and generalization. The integration of feature engineering, class imbalance handling, dimensionality reduction, and hyperparameter turned significantly improved predictive performance.

The local dataset provided granular categorization of blood pressure levels, allowing for more clinically relevant predictions. In contrast, the global dataset, though initially binary, was successfully adapted into a multi-class setting to enhance comparability. The performance gap between simple models (Logistic Regression) and ensemble methods (Random Forest, XGBoost) highlighted the importance of selecting advanced algorithms for healthcare prediction tasks.

Overall, the study established that predictive modeling is a viable and effective approach to early hypertension detection. The models developed can be integrated into clinical decision support systems or patient-facing applications to enhance patient safety, encourage preventive measures, and reduce the long-term burden of cardiovascular complications. Further research should emphasize the collection of larger and more diverse local datasets to improve generalizability, as local datasets reflect region-specific health patterns that may differ from global data.

Also, future research should integrate non-traditional data sources such as wearable devices (such as smartwatches, fitness trackers) with variables that may just offer additional predictive power and help capture real-time health dynamics beyond standard clinical data.

# 6. REFERENCES

[1] WHO (2023). Hypertension. World Health Organization (WHO). https://www.who.int/news-room/fact-sheets/detail/hypertension

[2] Vidal-Petiot, E. (2022). Thresholds for hypertension definition, treatment initiation, and treatment targets: Recent Guidelines at a glance. Circulation, 146(11), 805–807. https://doi.org/10.1161/circulationaha.121.055177

[3] World Heart Federation. (2024). World Heart Report 2023: Full Report - World Heart Federation. https://world-heart-federation.org/resource/world-heart-report-2023/

[4] Montagna, S., Pengo, M. F., Ferretti, S., Borghi, C., Ferri, C., Grassi, G., Muiesan, M. L., & Parati, G. (2022). Machine Learning in Hypertension Detection: A study on World Hypertension Day data. Journal of Medical Systems, 47(1). https://doi.org/10.1007/s10916-022-01900-5

[5] Haruna, I.S., Abubakar, S.S., Ibrahim, A., Osi, A.A., & 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

[6] Mondal, A. (2024). Predictive Modeling for Hypertension Detection: a Machine Learning approach. IJNRD.org. https://ijnrd.org/viewpaperforall.php?paper=IJNRD2405537

[7] Obafemi, A. S. (2022). A predictive model for predicting blood pressure levels using machine learning techniques. https://norma.ncirl.ie/6241

[8] Effati, S., Kamarzardi-Torghabe, A., Azizi-Froutaghe, F., Atighi, I., & Ghiasi-Hafez, S. (2024). Web application using machine learning to predict cardiovascular disease and hypertension in mine workers. Scientific Reports, 14, 31662. https://doi.org/10.1038/s41598-024-80919-9

[9] Kurniawan, R., Utomo, B., Siregar, K. N., Ramli, K., Besral, B., Suhatril, R. J., & Pratiwi, O. A. (2023). Hypertension prediction using machine learning algorithm among Indonesian adults. IAES International Journal of Artificial Intelligence (IJ-AI), 12(2), 776. https://doi.org/10.11591/ijai.v12.i2.pp776-784

[10] Garg, A., Sharma, B., & Khan, R. (2021). Heart disease prediction using machine learning techniques. IOP Conference Series. Materials Science and Engineering, 1022(1), 012046. https://doi.org/10.1088/1757-899x/1022/1/012046

[11] Islam, M., Alam, J., Maniruzzaman, M., Ahmed, N.A.M.F., Ali, S., Rahman, J., & Roy, D.C. (2023). Predicting the risk of hypertension using machine learning algorithms: A cross-sectional study in Ethiopia. PLoS ONE, 18(8), e0289613. https://doi.org/10.1371/journal.pone.0289613

[12] Jeong, Y. W., Jung, Y., Jeong, H., Huh, J. H., Sung, K.-C., Shin, J.-H., Kim, H. C., Kim, J. Y., & Kang, D. R. (2022). Prediction Model for Hypertension and Diabetes Mellitus Using Korean Public Health Examination Data (2002–2017). Diagnostics, 12, 1967. https://doi.org/10.3390/diagnostics12081967

[13] Sandhiya, R., Tharun, D., Subhika, S., Suriya, K. K., Sruthi, K., & Madhumitha, S. (2024). Enhanced Hypertension Disease Prediction using Machine Learning. Proceedings of the 3rd International Conference on Optimization Techniques in the Field of Engineering (ICOFE-2024). SSRN: https://ssrn.com/abstract=5091273 or http://dx.doi.org/10.2139/ssrn.5091273

[14] Das, S., Roy, S. K., Chakraborty, S., Dey, P., & Ghosh, D. (2024). Prediction of blood pressure using machine learning. International Journal of Intelligent Systems and Applications in Engineering, 12(22s), 579–586.

[15] Du, J., Chang, X., Ye, C., Zeng, Y., Yang, S., Wu, S., & Li, L. (2023). Developing a hypertension visualization risk prediction system utilizing machine learning and health check-up data. Scientific Reports, 13(1). https://doi.org/10.1038/s41598-023-46281-y

[16] Hwang, S. H., Lee, H., Lee, J. H., Lee, M., Koyanagi, A., Smith, L., Rhee, S. Y., Yon, D. K., & Lee, J. (2024). Machine Learning–Based Prediction for Incident Hypertension based on regular health checkup data: Derivation and validation in 2 independent nationwide cohorts in South Korea and Japan. Journal of Medical Internet Research, 26, e52794. https://doi.org/10.2196/52794

[17] Wanriko, S., Hnoohom, N., Wongpatikaseree, K., Jitpattanakul, A., & Musigavong, O. (2021). Risk Assessment of Pregnancy-induced Hypertension Using a Machine Learning Approach. 2021 Joint 6th International Conference on Digital Arts, Media and Technology with 4th ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunication

[18] Chang, W., Liu, Y., Xiao, Y., Yuan, X., Xu, X., Zhang, S., & Zhou, S. (2019). A Machine-Learning-Based prediction method for hypertension outcomes based on medical data. Diagnostics, 9(4), 178. https://doi.org/10.3390/diagnostics9040178

[19] Ishima, M.D., & Ikirigo, S.A. (2024). Detection and Classification of Malicious Websites Using Natural Language Processing (NLP) and Machine Learning (ML) Techniques. International Journal of Scientific Research in Science & Engineering Technology (IJSRSET), 11(6), 206–221. https://doi.org/10.32628/IJSRSET2411449

[20] Hasannejad, A. (2022). Data preprocessing in machine learning. Kaggle. https://www.kaggle.com/code/alirezahasannejad/data-preprocessing-in-machine-learning

[21] Ippolito, P. P. (2019). Feature engineering techniques. https://ppiconsulting.dev/blog/blog30/

[22] Jurafsky, D., & Martin, J. H. (2025). *Speech and Language Processing.* Draft of January 12, 2025. https://web.stanford.edu/~jurafsky/slp3/5.pdf

[23] Banoula, M. (2025). An Introduction to logistic regression in Machine learning. Simplilearn.com. https://www.simplilearn.com/tutorials/machine-learning-tutorial/logistic-regression-in-python

[24] Sakshi, A. (2023). Machine Learning algorithms, perspectives, and real-world application: Empirical evidence from United States trade data. *Indian Institute of Foreign Trade.* MPRA Paper No. 116579. https://mpra.ub.uni-muenchen.de/116579/

[25] Kavlakoglu, E., & Russi, E. (2025, May 19). XGBoost. *IBM.* https://www.ibm.com/think/topics/xgboost

[26] Tyagi, A. (2025). What is XGBoost Algorithm? *Analytics Vidhya.* https://www.analyticsvidhya.com/blog/2018/09/an-end-to-end-guide-to-understand-the-math-behind-xgboost/

[27] Eldem, A. (2025). A new hybrid learning model for early diagnosis of hypertension using IoMT technologies. *Ain Shams Engineering Journal (ASEJ).* https://doi.org/10.1016/j.asej.2025.103490

[28] Srivastava, T. (2025). 12 Important model evaluation Metrics for Machine Learning Everyone should know (Updated 2025). Analytics Vidhya. https://www.analyticsvidhya.com/blog/2019/08/11-important-model-evaluation-error-metrics/

[29] Schlosser, T., Friedrich M., Meyer, T., and Kowerko, D (2024). A Consolidated Overview of Evaluation and Performance Metrics for Machine Learning and Computer Vision. Junior Professorship of Media Computing, Chemnitz University of Technology, 09107 Chemnitz, Germany. https://www.researchgate.net/publication/374558675_A_Consolidated_Overview_of_Evaluation_and_Performance_Metrics_for_Machine_Learning_and_Computer_Vision

[30] Swaminathan, Sathyanarayanan & Tantri, B Roopashri. (2024). Confusion Matrix-Based Performance Evaluation Metrics. African Journal of Biomedical Research. 27. 4023-4031. 10.53555/AJBR.v27i4S.4345.

[31] Narkhede, S. (2018). Understanding AUC - ROC Curve. https://medium.com/data-science/understanding-auc-roc-curve-68b2303cc9c5