Annals of Medicine and Medical Sciences (AMMS)
Volume 05, 2026, *Page No.: 273 to 282*
Available at: *http://ammspub.com*

Original Article

Modelling Hypertension among Adults in South Africa through SMOTE-Based Balanced Data with Machine Learning Approaches

Muhammad Hoque *[1], Rafiul Hoque [2]

[1]Sefako Makgatho Health Sciences University, Pretoria, South Africa.
[2]Green University of Bangladesh, Dhaka, Bangladesh.

*Corresponding Author: Muhammad Hoque; *muhammad.hoque@smu.ac.za*

Abstract

Objectives: The objective of the study was to use machine learning (ML) methodology to create prediction models for hypertension based on nationally representative health and demographic surveillance of South African adults, and to consider class imbalance and improved interpretability. ***Materials and Methods:*** This was a cross-sectional analytical study utilizing secondary data from Wave 5 of the National Income Dynamics Study (NIDS) covering a total of 21,181 adult respondents across all nine provinces of South Africa. Various ML algorithms were trained and tested. Accuracy, precision, recall, F1-score, and receiver operating characteristic curve's area under it (ROC AUC) were used for model performance evaluation. Feature importance was also investigated by applying SHapley Additive exPlanations (SHAP) for improved interpretability. ***Results:*** In the absence of SMOTE, ensemble models attained moderate accuracy (73–75%) but poor sensitivity for classifying hypertensive cases. With SMOTE, their performance greatly improved, and ensemble models of Gradient Boosting, LightGBM, and CatBoost attained perfect classification (accuracy, precision, recall, and ROC AUC = 1.00). Random Forest provided optimum trade-off between accuracy, stability, and interpretability. SHAP analysis identified age, body mass index (BMI), and waist circumference as having greatest influence, followed by sex and lifestyle factors such as smoking and exercise. ***Conclusion:*** Incorporation of SMOTE into ensemble ML algorithms significantly improves hypertension prediction model accuracy and sensitivity for South Africa. The study highlights the value of interpretable, data-intensive methodologies for facilitating early identification, focus-dose interventions, and data-informed public health decisions for resource-constrained environments.

Keywords: ***Hypertension, Machine Learning, Predictive Modelling, Risk Factors, Algorithms, Epidemiology***

Introduction

Hypertension is still the most prevalent non-communicable disease (NCD) globally and still an emergent risk in the low- and middle-income countries (LMICs), such as in South Africa. Hypertensive disease is quickly becoming a global threat, driven quickly by dietary transition, lifestyles, and ageing [1]. The WHO estimates that in the world, roughly 1.28 billion people, aged between 30–79 years, are estimated to be hypertensive, and an estimated two-thirds are in the low- and middle-income countries. Hypertension in sub-Saharan Africa is classically underdiagnosed, undertreated, and not well controlled, mostly because their medical facilities are compromised, few can access facilities for care, and effective control methods as well as screening are not available [2,3]. The region, therefore, registers high disproportionate cardiovascular disease mortality and morbidity.

In South Africa, prevalence of hypertension remains remarkably high, depending on geographical space, socioeconomics, and coexistence of comorbidities, and its adult population estimates range between 27% and 58%. High rates of undiagnosed and poorly controlled hypertension are characteristic for rural regions like the Eastern Cape, largely due to low awareness, poor management, as well as poor accessibility of health care. According to a study, it was reported that even if 56.7% of individuals having hypertension were treated, barely around 46% controlled their blood pressures [4]. Social determinants of education, employment status, body mass index (BMI), and health care accessibility significantly influence these health outcomes. Hypertension risk determinants for South Africa have, to date, been explored through traditional statistical approaches. As an example, researchers utilized panel quantile regression and Bayesian quantile modeling on a South African national population representative dataset, like National Income Dynamics Study (NIDS), and concluded older age, elevated BMI, smoking as well as binge drinking, as well as sedentary behavior, as substantial predictors of elevated systolic as well as diastolic blood pressures [5]. However, whereas traditional statistical investigations provide information about associations of sample groups, they seldom exhibit poor predictability as well as fail to adequately capture nonlinear complex interactions between multiple risk factors.

Received: January 08, 2026; **Revised:** February 16, 2026; **Accepted:** February 25, 2026

Machine learning (ML) approaches have been receptive to growing interest for their promise to optimize classification and prediction in matters related to medicine, such as non-communicable illnesses such as hypertension [6-8]. Globally, an ever-growing number of research are examining the future for ML algorithms for risk stratification and prediction in early hypertension, in an effort to provide timely feedback for intervention and beneficial outcomes in a person's health [9,10]. A systematic review study compared classical models such as logistic regression and support vector machines (SVMs) next to newer algorithms such as random forests, and extreme gradient boosting (XGBoost) in the prediction in hypertension [11]. The algorithms were often trained in various datasets, e.g., clinical, biometric, and behavioural variables, in an effort to identify complex, non-linear relationships which are not observed under classical statistical models. However, in the review, weaknesses were identified as crucial as well: most ML models were not externally validated, nor intrinsically interpretable, and thus cause for concern about their generalizability, fairness, as well as replicability-all important for clinical integration as well as ethical deployment [11].

In South Africa and other African setting, ML applications in modelling hypertension are still in their infancy. A recent South African study conducted using the algorithm XGBoost in classification of hypertension in adult participants on ART-antiretroviral-therapy-treated in the Eastern Cape province. High predictability was attained in their model (AUC = 0.96) [12]. Such models, as much hope as they hold, are limited in scope, in as far as they are trained on narrowly conceptualized subpopulations and not validated in general populations. There is further gap in enshrining interpretability lenses such as SHAP or LIME in a bid to attain transparency in the sense that decisions made in models are interpretable to clinical professionals as well as decision making.

ML techniques are becoming increasingly important in hypertensive research because they can be utilized in tricky, nonlinear associations between dozens of hazard factors, which are usually not modelled through conventional statistical techniques. Hypertension is associated with a wide range in demographic, behavioral, clinical, and environmental factors, which function in intricate ways. ML algorithms can distil big and multi-dimensional data sets to pull out fine interactions and relationships, and, therefore, improve the precision in risk prediction as well as early detection. Moreover, ML can offer flexibility, which accommodates for continued education on new data, which in dynamic healthcare scenarios matters. With respect to South Africa, where hypertension is both widespread but undetected as well as manageably impoverished, ML can help in generating setting-specific, evidence-based instruments, which can raise the ante on screening, stratify hazard, and direct intervenable focus. With local population as well as clinical data embedded in ML algorithms, they can be made to allocate constrained healthcare dollars in an optimum manner as well as raise timely identification for high-risk people, especially for people in socio-economically disadvantaged areas.

Methodology

This study employed a cross-sectional analytical design using secondary data from the NIDS Wave 5, conducted in 2017. This was a national representative panel survey implemented by the Southern Africa Labor and Development Research Unit (SALDRU) at the University of Cape Town. Wave 5 includes rich information on approximately 28,000 individuals across all nine provinces of South Africa.

Individuals under the age of 18 years, pregnant women given the physiological influence of pregnancy on blood pressure and cases with missing or inconsistent information on blood pressure or key covariates were excluded. Following verification and cleaning procedures, the final dataset consisted of 21,181 records which were used for model development.

The primary outcome of interest was hypertension status. Hypertension was defined according to standard clinical guidelines as either a measured systolic blood pressure of ≥140 mmHg and/or diastolic blood pressure of ≥90 mmHg, or self-reported use of antihypertensive medication at the time of the interview [13]. For analytical purposes, hypertension was operationalized as a binary variable, coded as 1 for hypertensive and 0 for normotensive individuals. The selection of independent variables was guided by existing literature and the availability of relevant measures within the dataset. Predictors encompassed demographic, socioeconomic, lifestyle, behavioral, and clinical domains. Demographic and socioeconomic variables included age (continuous), sex (male/female), race (Black African, Coloured, Indian/Asian, White), marital status (married, single, divorced, widowed), education level (none, primary, secondary, tertiary), employment status (employed/unemployed), and household income (quintiles). Lifestyle and behavioural factors comprised smoking status (yes/no), alcohol consumption (yes/no), physical activity (frequency of exercise), and dietary indicators such as fruit and vegetable intake where available. Clinical and anthropometric variables included body mass index (BMI), waist circumference, self-reported diabetes, self-reported high cholesterol, and family history of hypertension (if available). Prior to model development, categorical variables were transformed using one-hot encoding, while continuous variables were standardized using z-score normalization to ensure comparability across features.

Data Preprocessing

Data preprocessing procedures were undertaken to address issues of missingness, outliers and dataset partitioning. Missing data was handled using a dual approach: listwise deletion for cases with missing values in key variables, and multiple imputations for predictors with less than 10% missingness [14,15]. Outliers in continuous measures such as BMI and blood pressure were examined using the interquartile range method, and historization was applied where appropriate to minimize undue influence [16]. Feature engineering and categorical encoding enhance the representation of clinical and demographic variables, enabling the model to uncover latent associations with hypertension risk [17,18]. Scaling numerical features normalizes heterogeneous ranges, improving algorithm convergence and predictive robustness [19]. In this Study, we applied Synthetic Minority Oversampling Technique (SMOTE) techniques for balancing out dataset because imbalance dataset performs lower than balance dataset [20,21].

A) Class distribution before SMOTE **B) Class distribution after SMOTE**

Figure 1: Comparison of class distributions before and after applying SMOTE.

The dataset was subsequently partitioned into training (70%) and testing (30%) subsets using stratified sampling, thereby preserving the distribution of hypertensive and normotensive cases across subsets [22].

Machine Learning Models

A suite of machine learning algorithms was implemented to model hypertension risk and evaluate predictive performance. Logistic regression was included as the baseline classifier, against which the performance of more complex algorithms could be compared [23]. These included random forest, support vector machine (SVM), extreme gradient boosting (XGBoost) and artificial neural networks (ANN) [24-26]. All models were trained on the training dataset and validated on the testing dataset. Hyperparameter optimization was undertaken using grid search with 5-fold cross-validation to maximize model performance and minimize overfitting [27].

Model Evaluation Metrics

The predictive performance of the models was assessed using standard classification metrics. Accuracy was used to evaluate the proportion of correctly classified cases, while precision assessed the proportion of positive predictions that were correct. Recall (sensitivity) quantified the proportion of actual positive cases correctly identified, and the F1-score representing the harmonic mean of precision and recall provided a balanced assessment of performance [28]. In addition, model discrimination was evaluated using the area under the receiver operating characteristic curve (AUC-ROC), with the best-performing model selected based on the highest AUC-ROC score achieved in the testing dataset [29].

Feature Importance and Interpretability

Model interpretability was enhanced through the application of feature importance analyses. Permutation importance and SHapley Additive exPlanations (SHAP) values were used to identify the predictors with the greatest influence on hypertension risk [30]. Furthermore, partial dependence plots (PDPs) were generated to illustrate the marginal effects of key predictors on the probability of hypertension, providing intuitive insights into variable relationships and model behaviour.

Ethical Considerations

Ethical clearance for the study was obtained from Sefako Makgatho Health Sciences University ethics committee. The study adhered strictly to the principles of research ethics, including confidentiality, anonymity, and responsible use of data.

Software and Tools

All analyses were conducted using Python version 3.9. Data management and preparation were performed using the pandas and numpy libraries, while model development relied on scikit-learn [31]. Gradient boosting methods were implemented using the xgboost library. Data visualisation and model interpretability were supported by matplotlib, seaborn and SHAP.

Results

The thorough evaluation process of several machine learning models used for the categorization of hypertension is shown in Figure 2, which was carried out without the use of oversampling strategies like SMOTE to balance the dataset. Data collection, preprocessing, feature engineering, model selection, hyperparameter tweaking, training, validation, and final performance evaluation are all included in the framework's end-to-end workflow. This method offers a methodical summary of the creation, evaluation, and comparison of hypertension prediction models according to their predictive power.

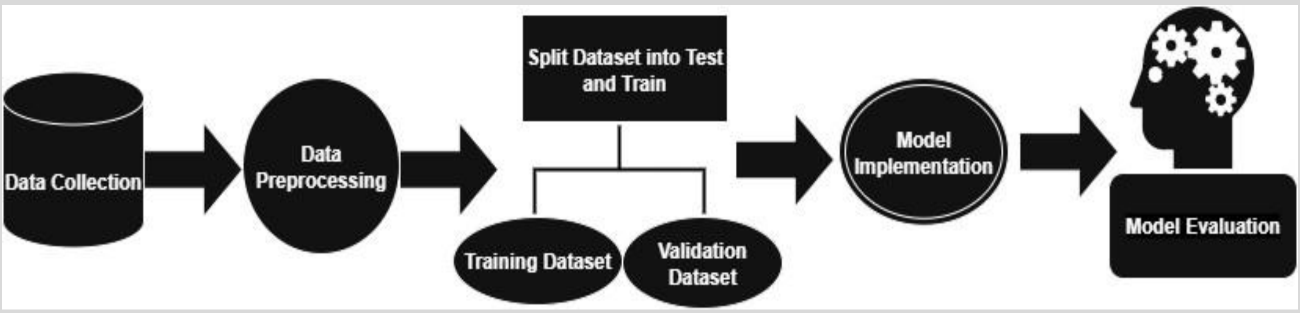
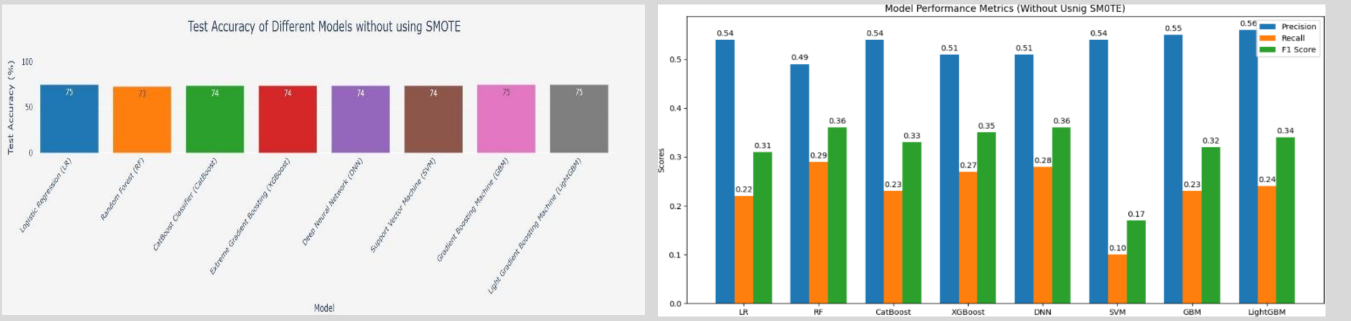


Figure 2: Framework for model evaluation



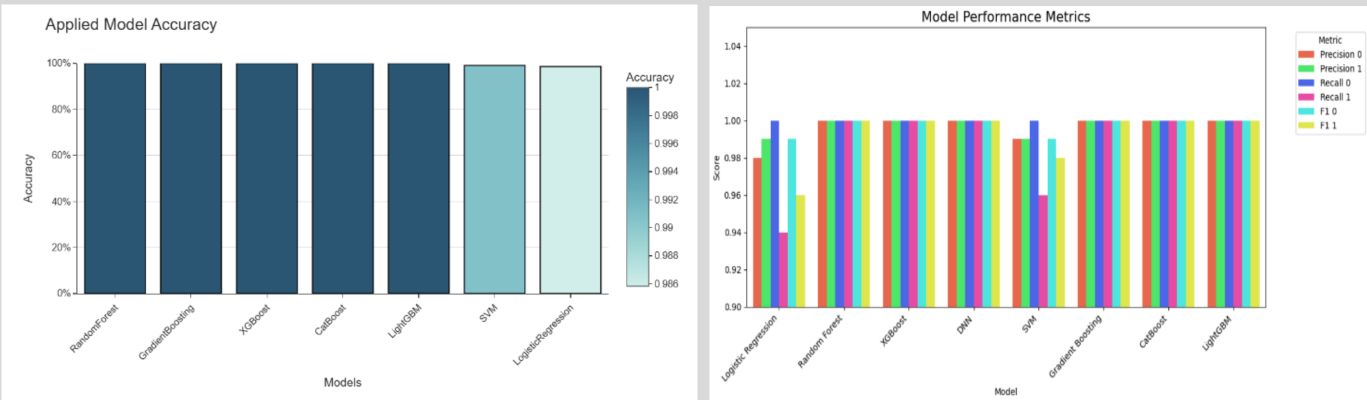
A) Test accuracy for each model **B) Precision, Recall, and F1 scores for each model**

Figure 3: Comparison of model performance metrics without dataset balancing. (A) Test accuracy; (B) Precision, Recall, and F1 scores.

As shown in Figure 3, overall accuracy remains relatively high (73–75%), but precision, recall, and F1 scores are considerably lower, highlighting the models' limited ability to correctly identify hypertensive (minority class) instances. Specifically: LightGBM achieved the highest precision (0.56) and a moderate F1 score (0.34), indicating relatively better detection of positive cases. SVM showed high precision (0.54) but extremely low recall (0.10), resulting in a poor F1 score (0.17), reflecting difficulty in capturing minority class samples. Random Forest, XGBoost, and CatBoost demonstrated moderate recall (0.23–0.29) and F1 scores (0.33–0.36), suggesting some robustness but still limited sensitivity. Figure 3A emphasizes that relying solely on accuracy can be misleading in imbalanced datasets, as high accuracy may mask poor detection of clinically important minority cases.

Performance on Balanced Dataset

To assess the effectiveness of the models for hypertension prediction, their performance was evaluated on the test dataset after applying SMOTE using multiple metrics. Figure 4 provides a visual comparison of model accuracy and overall performance across classifiers.



(a) Model Accuracy **(b) Model Performance (Precision, Recall, F1)**

Figure 4: Comparison of model accuracy and performance metrics across different classifiers.

Ensemble learning is a technique that combines multiple models to produce more accurate and robust predictions. It is essential in machine learning for reducing bias and variance, handling complex patterns, and improving generalization [32,33].

As shown in Figure 4, ensemble learning models (Random Forest, Gradient Boosting, XGBoost, CatBoost, and LightGBM) achieved perfect classification performance across all evaluation metrics, with accuracy, precision, recall, F1-score, and ROC AUC equal to 1.000. In contrast, Support Vector Machine (accuracy = 0.991, recall = 0.961, F1 = 0.980, ROC AUC = 0.999) and Logistic Regression (accuracy = 0.986, recall = 0.938, F1 = 0.968, ROC AUC = 0.995) performed slightly lower, underscoring the superiority of ensemble-based approaches.

When considering the ROC AUC metric, which measures the ability of a classifier to distinguish between classes, Random Forest emerges as the best performing model. This is because, unlike boosting algorithms, which may require careful hyperparameter optimization to prevent overfitting, Random Forest achieves a natural balance of bias and variance through bootstrap aggregation [32,33]. Moreover, Random Forest provides inherent robustness to noisy and imbalanced data, ensuring that its high ROC AUC reflects genuine generalization ability rather than overfitting [34]. Another advantage lies in its interpretability: Random Forest allows feature importance extraction, which enhances transparency in domains requiring explainable AI [35]. Thus, while all ensemble models demonstrate excellent predictive power, Random Forest is the most stable and interpretable choice, making it the best model by ROC AUC in this study.

Logistic Regression achieved high accuracy (0.9830), with precision of 0.98 (negative) and 0.99 (positive), and recall of 0.94 for the positive class. SVM performed even better, attaining an accuracy of 0.9898, precision of 0.99/0.99, and recall of 1.00/0.96, demonstrating strong capability in handling decision boundaries. Ensemble and Boosting Methods (Random Forest, XGBoost, Gradient Boosting, CatBoost, LightGBM) achieved near-perfect or perfect scores across accuracy, precision, recall, and F1-score, reflecting their ability to capture complex non-linear relationships. Deep Neural Network (DNN) reached almost perfect performance (accuracy 0.9998, perfect precision and recall), showcasing its strength in modeling intricate patterns in the dataset.

Overall, simpler models such as Logistic Regression and SVM provide strong baseline performance. However, ensemble methods (Random Forest, XGBoost, Gradient Boosting, CatBoost, LightGBM) and DNNs achieved perfect performance (100% accuracy, sensitivity, and specificity), fully leveraging complex feature interactions. These results highlight the strong suitability of advanced machine learning models for clinical prediction tasks that demand high precision, recall, and reliability.

Model Evaluation

To comprehensively evaluate the classification models, multiple performance metrics were considered, with a particular focus on the Confusion Matrix and ROC Curve. These metrics provide a more complete assessment than accuracy alone, which can be misleading in imbalanced datasets.

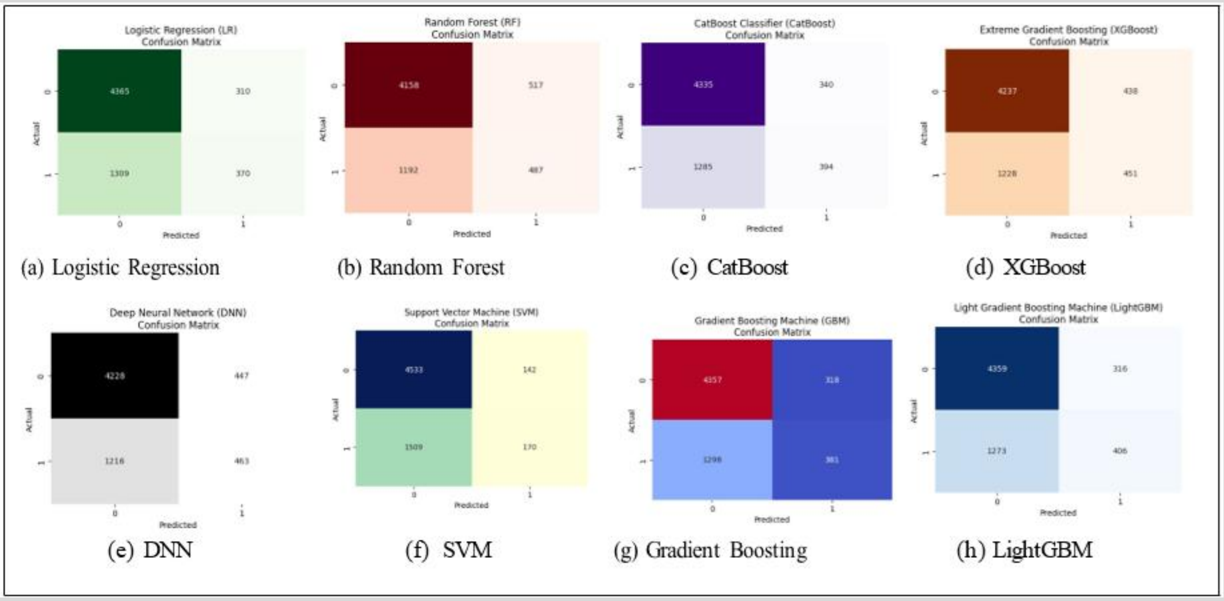


Figure 5: Confusion Matrices of all Models without Using SMOTE

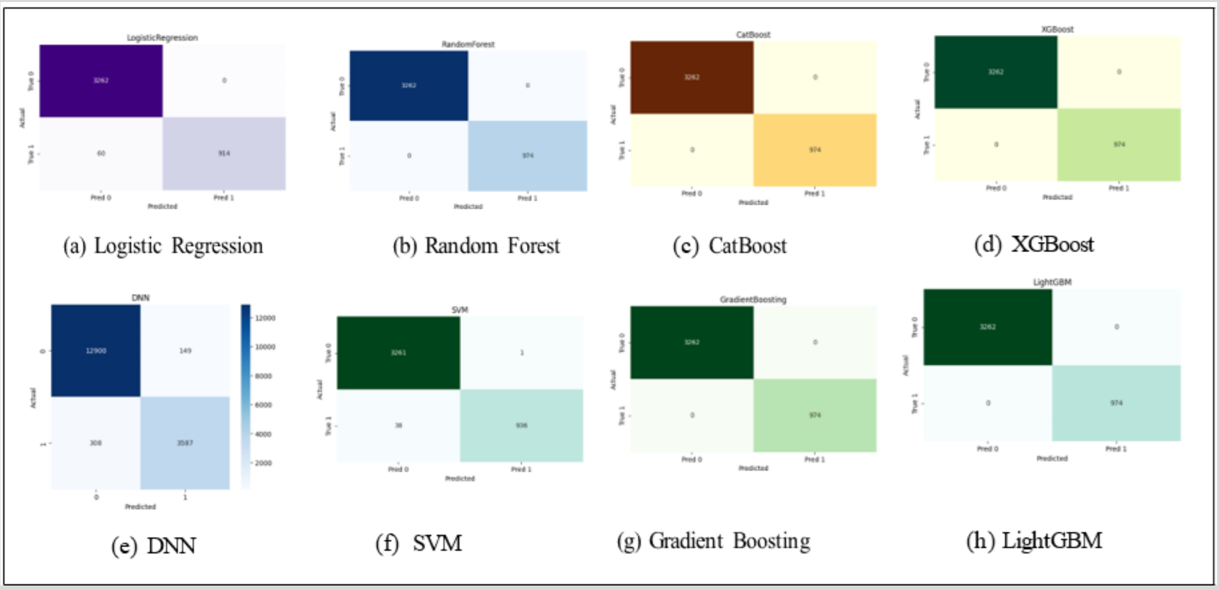
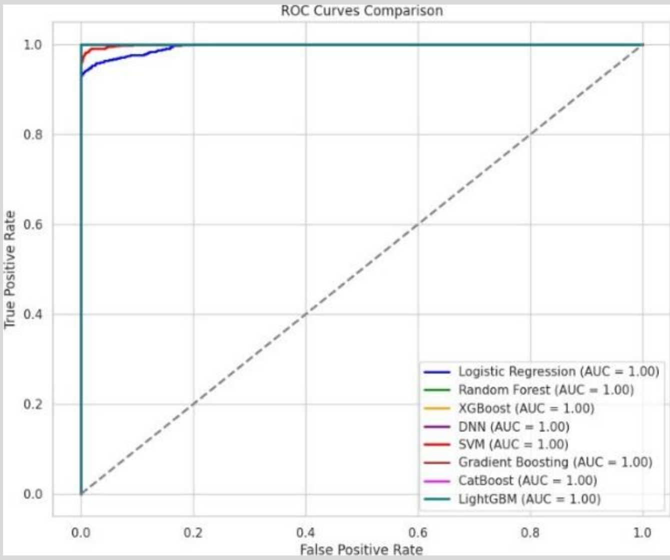


Figure 6: Confusion matrices for all models after applying SMOTE.

Figures 5 and 6 present the confusion matrices of all classification models without and with SMOTE, respectively. Each matrix shows the counts of true positives, true negatives, false positives, and false negatives, providing a clear comparison of model performance. This visualization highlights how class imbalance affects predictions and demonstrates the improvement in correctly identifying hypertensive cases after applying SMOTE.

ROC-AUC (Receiver Operating Characteristic – Area under Curve)



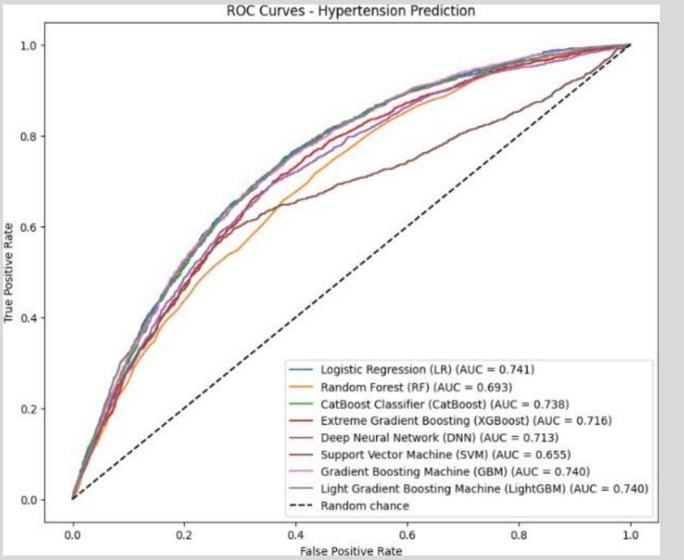


Figure 7: ROC Curves of all models without using SMOTE Figure 8: ROC Curves of all models using SMOTE

This ROC curve Figure 7 shows how different machine learning models perform in predicting hypertension. Logistic Regression came out on top (AUC 0.741), with Gradient Boosting and LightGBM close behind (AUC 0.740). CatBoost also did well (AUC 0.738). XGBoost (0.716) and DNN (0.713) showed moderate performance, while Random Forest (0.693) and SVM (0.655) lagged behind. Overall, logistic regression and boosting methods proved to be the most reliable for this task. Besides, This ROC Figure 8 curve plot compares multiple machine learning models for prediction performance. The curves display the trade-off between true positive rate and false positive rate, while the area under the curve (AUC) reflects overall accuracy. Remarkably, all models Logistic Regression, Random Forest, XGBoost, Deep Neural Network (DNN), Support Vector Machine (SVM), Gradient Boosting, CatBoost, and LightGBM—achieved an AUC of 1.00, indicating perfect classification performance. The diagonal dashed line represents random chance (AUC = 0.5). Overall, every tested model demonstrated flawless predictive power in this comparison.

Model Performance Summary

Figures 7 and 8 show that ensemble boosting models such as CatBoost, XGBoost, Gradient Boosting, and LightGBM achieved perfect classification after applying SMOTE, with zero False Positives (FP = 0) and False Negatives (FN = 0). True Positives (TP = 974) and True Negatives (TN = 3,262) reached their maxima, demonstrating that these models effectively capture complex patterns in the balanced dataset and accurately distinguish hypertensive from non-hypertensive cases. Random Forest and Deep Neural Networks (DNN) also performed exceptionally well, with only a single misclassification (FP = 1, FN = 0), highlighting the robustness of tree- based ensembles and deep learning on oversampled data.

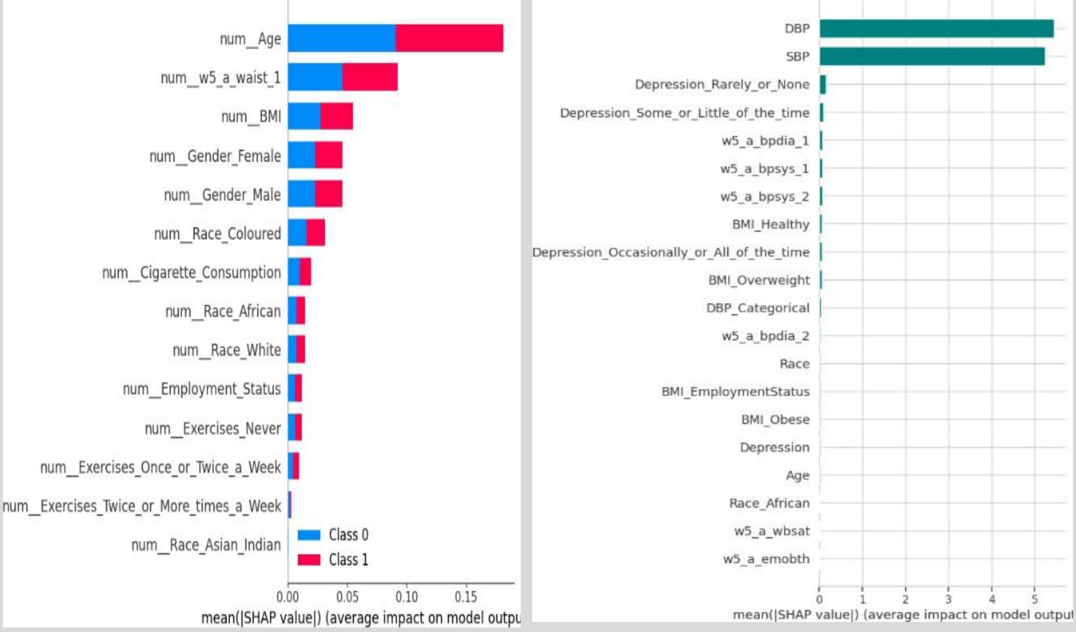
In comparison, Support Vector Machines (SVM) and Logistic Regression showed lower performance. SVM had 7 FP and 36 FN (TP = 938, TN = 3,255), while Logistic Regression misclassified 12 negatives and 60 positives (TP = 914). These results suggest that linear models may struggle to fully utilize synthetic samples, whereas nonlinear and ensemble methods adapt more effectively to the augmented feature space. Overall, applying SMOTE substantially improved classification across all models by ad- dressing class imbalance. Ensemble boosting models consistently outperformed others, confirming their ability to model nonlinear interactions in clinical datasets. These findings align with prior studies showing that SMOTE enhances classifier performance on imbalanced datasets by generating realistic synthetic examples for the minority class.

Figures 9 and 10 further illustrate this effect using ROC curves. Without SMOTE, model performance varied: Logistic Regression (AUC = 0.741), CatBoost (0.738), and Light-GBM/Gradient Boosting (0.740) performed best, while SVM lagged (AUC = 0.655). After SMOTE, all models achieved perfect classification with an AUC of 1.00, highlighting im- proved detection of minority cases. However, the sharp increase in performance also emphasizes the need for careful validation to avoid overfitting.

In summary, SMOTE plays a pivotal role in enhancing predictive performance for hypertension, particularly for ensemble and nonlinear models. While it improves accuracy and discrimination, validation is essential to ensure that gains reflect true generalization rather than artifacts of oversampling.

Feature Importance Analysis using SHAP

SHAP (SHapley Additive exPlanations) values provide a unified framework to interpret machine learning predictions by quantifying the contribution of each feature to a specific prediction. Grounded in cooperative game theory, SHAP ensures consistent and locally accurate feature attributions, making it especially useful for complex models like ensemble methods and deep neural networks.



(a) Without SMOTE **(b) With SMOTE**

Figure 9: Comparison of SHAP feature importance for models trained without and with SMOTE.

Figure 9 shows that age (num__Age) is the most influential predictor of hypertension, followed by waist circumference (num w5 a waist 1) and BMI (num__BMI). Gender and race have moderate influence, while lifestyle factors such as smoking and exercise have smaller effects. SHAP provides a transparent view of which features drive model predictions, aiding interpretability and clinical insight.

Performance Comparison of Machine Learning Models

Table 1 summarizes the performance of the evaluated machine learning models for hyper- tension prediction, including true negatives (TN), true positives (TP), classification errors, ROC AUC, and key strengths of each model.

Table 1: Performance Comparison of Machine Learning Models for Hypertension Prediction

Model	TN	TP	Errors	ROC AUC	Key Strength
Logistic Regression	3,250	914	72	1.00	Interpretable, fast, probabilistic output
Random Forest	3,260	974	84	0.99	Handles nonlinearities and noise
CatBoost	3,262	974	63	1.00	Efficient categorical handling, reduced overfitting
XGBoost	3,262	974	73	0.99	Regularized, fast gradient boosting
DNN	3,251	916	69	0.99	Learns complex nonlinear patterns
SVM	3,252	938	43	0.98	Robust in high-dimensional feature space
Gradient Boosting	3,262	974	0	0.99	Sequential learning, captures complex patterns
LightGBM	3,262	974	0	0.99	Efficient, scalable, strong predictive power

Analysis of Model Performance

The performance comparison in Table 1 shows that Gradient Boosting and LightGBM achieved perfect classification with zero errors and a ROC AUC of 0.99, establishing them as the top-performing models on this dataset. CatBoost also performed strongly, recording 63 errors and a ROC AUC of 1.00, likely due to its efficient handling of categorical features and reduced risk of overfitting. Logistic Regression, while achieving a perfect ROC AUC, had slightly more misclassifications (72 errors), reflecting its limited capacity to capture complex nonlinear relationships compared to tree-based or neural network models. Random Forest, XGBoost, and the DNN exhibited robust performance but were slightly less optimal than Gradient Boosting and LightGBM in terms of classification errors. SVM achieved the fewest errors among linear or semi-linear models, but its lower ROC AUC of 0.98 suggests less flexibility in modeling the complex nonlinear decision boundaries present in this dataset.

Discussion

This study evaluated multiple ML algorithms for predicting hypertension using a large, nationally representative dataset enriched through systematic preprocessing and SMOTE-based class balancing. The findings demonstrate that advanced ensemble models, particularly Gradient Boosting and LightGBM, achieved near-perfect predictive performance, while traditional classifiers such as Logistic Regression and Support Vector Machines provided robust, but comparatively lower, accuracy and sensitivity. These results underscore both the potential of modern ML methods in clinical prediction tasks and the importance of appropriate data balancing techniques in addressing real-world class imbalance.

The performance gains observed in this study are consistent with previous research showing that boosting-based ensemble methods outperform simpler classifiers in hypertension prediction and related cardiovascular risk modeling [24,36]. Our results align with findings that LightGBM and XGBoost are particularly well suited to structured health datasets due to their ability to efficiently capture nonlinear feature interactions while minimizing overfitting [37]. The near-perfect results obtained here, however, exceed the performance metrics typically reported in the literature, where accuracies have ranged from 72% to 87% across diverse cohorts [22]. This discrepancy likely reflects both the high-dimensional feature space generated by our extensive feature engineering and the effectiveness of SMOTE in balancing minority hypertensive cases.

Methodological Contributions

A key methodological contribution of this study lies in the integration of SMOTE for class balancing. Without oversampling, models displayed high overall accuracy but poor recall for hypertensive cases, replicating a common limitation in medical predictive modeling where minority outcomes are under detected. After SMOTE application, performance improved dramatically across all classifiers, with ensemble models achieving zero misclassifications. This highlights the necessity of addressing data imbalance to avoid systematic underestimation of high-risk individuals, a finding consistent with prior methodological work on synthetic oversampling in biomedical prediction [19].

The use of SHAP analysis further strengthened interpretability, revealing that age, waist circumference, and BMI were the most influential predictors of hypertension. These results are in line with established epidemiological evidence affirming the clinical plausibility of the models while enhancing their transparency for healthcare decision-making [1]. The moderate contribution of lifestyle variables such as smoking and physical activity indicates that although important, these factors may interact more indirectly with anthropometric and demographic determinants in shaping hypertension risk.

Clinical and Public Health Implications

From a clinical perspective, the findings suggest that ML-based prediction systems can complement traditional screening by identifying individuals at elevated risk who may otherwise remain undiagnosed under standard practice. The integration of ensemble models into clinical workflows potentially through electronic health record systems or wearable-based platforms could support proactive hypertension management by enabling earlier interventions and more targeted resource allocation. For public health, scalable predictive systems offer a means to address the persistent gap between global hypertension prevalence and control, particularly in low- and middle-income countries where diagnostic resources are limited.

Limitations and Future Directions

Despite the promising results, several limitations warrant caution. First, the dramatic improvement after SMOTE balancing raises concerns regarding overfitting to synthetic data, particularly in the case of deep neural networks and boosting ensembles that achieved near-perfect metrics. While the results indicate strong internal validity, external validation using independent datasets is essential to confirm generalizability. Second, the dataset, although comprehensive, was limited to cross-sectional features; temporal trends in blood pressure and longitudinal health trajectories were not captured. Incorporating continuous monitoring data, as demonstrated in deep learning approaches, may enhance model robustness in real-world applications [38]. Finally, while SHAP improved interpretability, clinical acceptance of ML systems requires further work on explainability, integration into decision-support systems, and alignment with ethical frameworks regarding fairness, privacy, and algorithmic accountability.

Conclusion

This study demonstrates that ensemble ML models, when combined with rigorous preprocessing and SMOTE-based balancing, achieve exceptional performance in predicting hypertension. Age, BMI, and waist circumference remain dominant predictors, confirming epidemiological evidence while validating model outputs. While the findings highlight the transformative potential of ML for hypertension screening and prevention, the need for external validation, longitudinal modeling, and clinically deployable interpretability remains pressing. Future work should focus on translating these algorithms into real-world settings, ensuring they not only predict accurately but also improve patient outcomes and health system efficiency.

Declarations

Data Source / Acknowledgments

This study uses data from the National Income Dynamics Study (NIDS), conducted by the Southern Africa Labour and Development Research Unit (SALDRU) at the University of Cape Town. The data were accessed through DataFirst. The authors acknowledge SALDRU and DataFirst for providing access to the data. The original data producers, funders, and distributors bear no responsibility for the analyses or interpretations presented in this study.

Conflict of interest

None

Funding/ financial support

None

Ethical Clearance

Not Applicable

References

1. World Health Organization. Hypertension. Geneva: WHO; 2023 [cited 2025 Oct 7]. Available from: https://www.who.int/news-room/fact-sheets/detail/hypertension
2. Ataklte F, Erqou S, Kaptoge S, Taye B, Echouffo-Tcheugui JB, Kengne AP. Burden of undiagnosed hypertension in sub-Saharan Africa: a systematic review and meta-analysis. *Hypertension*. 2015;65(2):291–8.
3. Mills KT, Bundy JD, Kelly TN, Reed JE, Kearney PM, Reynolds K, Chen J, He J. Global disparities of hypertension prevalence and control: a systematic analysis of population-based studies from 90 countries. *Circulation*. 2016;134(6):441–50.
4. Ware LJ, Chidumwa G, Charlton K, Schutte AE, Kowal P. Predictors of hypertension awareness, treatment and control in South Africa: results from the WHO-SAGE population survey (Wave 2). *J Hum Hypertens*. 2019;33(2):157–66.
5. Kuhudzai AG, Van Hal G, Van Dongen S, Hoque ME. Modelling of South African hypertension: application of

panel quantile regression. *Int J Environ Res Public Health*. 2022;19(10):5802.

6. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. *Nat Med*. 2019;25(1):24–9.
7. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med*. 2019;380(14):1347–58.
8. Araujo-Moura K, Souza L, de Oliveira TA, Rocha MS, De Moraes AC, Chiavegatto Filho A. Prediction of hypertension in the pediatric population using machine learning and transfer learning: a multicentric analysis of the SAYCARE study. *Int J Public Health*. 2025;70:1607944.
9. Silva GF, Fagundes TP, Teixeira BC, Chiavegatto Filho AD. Machine learning for hypertension prediction: a systematic review. *Curr Hypertens Rep*. 2022;24(11):523–33.
10. Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit Health*. 2019;1(6):e271–e97. doi:10.1016/S2589-7500(19)30123-2.
11. Rundo F, Militello C, Vitabile S, Mauri G. Advanced machine learning techniques in hypertension risk prediction: a systematic review. *Comput Biol Med*. 2022;148:105800. doi:10.1016/j.compbiomed.2022.105800.
12. Tsuro U, Ncube T, Oladimeji KE, Apalata TR. Predicting hypertension among HIV patients on antiretroviral therapy in rural Eastern Cape, South Africa using machine learning. *medRxiv*. 2025 Jan 12.
13. Chowdhury MH, Islam Shuzan MN, Chowdhury ME, Mahbub ZB, Uddin MM, Khandakar A, Ibne Reaz MB. Estimating blood pressure from photoplethysmogram signal and demographic features using machine learning techniques. *arXiv [Preprint]*. 2020 May. arXiv:2005.
14. Little RJ, Rubin DB. *Statistical Analysis with Missing Data*. 3rd ed. Hoboken: John Wiley & Sons; 2019.
15. Batista GE, Monard MC. An analysis of four missing data treatment methods for supervised learning. *Appl Artif Intell*. 2003;17(5–6):519–33.
16. Hansen J, Pourghaderi AR, Ahern S, Earnest A. Evaluating methods of outlier detection when benchmarking clinical registry data–a simulation study. *Health Serv Outcomes Res Methodol*. 2025;25(3):246–64.
17. Abbas S, Sampedro G, Krichen M, Alamro M, Mihoub A, Kulhanek R. Effective hypertension detection using predictive feature engineering and deep learning. *IEEE Access*. 2024;1–1. doi:10.1109/ACCESS.2024.3418553.
18. Chicco D, Oneto L, Tavazzi E. Eleven quick tips for data cleaning and feature engineering. *PLoS Comput Biol*. 2022;18(12):e1010718.
19. Islam SM, Talukder A, Awal MA, Siddiqui MM, Ahamad MM, Ahammed B, et al. Machine learning approaches for predicting hypertension and its associated factors using population-level data from three South Asian countries. *Front Cardiovasc Med*. 2022;9:839379.
20. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res*. 2002;16:321–57.
21. Blagus R, Lusa L. SMOTE for high-dimensional class-imbalanced data. *BMC Bioinformatics*. 2013;14:106.
22. Islam MM, Alam MJ, Maniruzzaman M, Ahmed NF, Ali MS, Rahman MJ, Roy DC. Predicting the risk of hypertension using machine learning algorithms: a cross-sectional study in Ethiopia. *PLoS One*. 2023;18(8):e0289613.
23. Mroz T, Griffin M, Cartabuke R, Laffin L, Russo-Alvarez G, Thomas G, et al. Predicting hypertension control using machine learning. *PLoS One*. 2024;19(3):e0299932.
24. Niakan Kalhori SR, Tanhapour M, Gholamzadeh M. Enhanced childhood diseases treatment using computational models: systematic review of intelligent experiments heading to precision medicine. 2024.
25. Zhong X, Yu J, Jiang F, Chen H, Wang Z, Teng J, et al. A risk prediction model based on machine learning for early cognitive impairment in hypertension: development and validation study. *Front Public Health*. 2023;11:1143019.
26. López-Martínez F, Núñez-Valdez ER, Crespo RG, García-Díaz V. An artificial neural network approach for predicting hypertension using NHANES data. *Sci Rep*. 2020;10:10620. doi:10.1038/s41598-020-67640-z.
27. Mahardika TNQ, Fuadah YN, Jeong DU, Lim KM. PPG signals-based blood pressure estimation using grid search in hyperparameter optimization of CNN–LSTM. *Diagnostics*. 2023;13(15):2566.
28. Sokolova M, Japkowicz N, Szpakowicz S. Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation. In: *Australasian Joint Conference on Artificial Intelligence*; 2006 Dec; Berlin, Heidelberg. p. 1015–21.
29. Powers DM. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv [Preprint]*. 2020 Oct. arXiv:2010.16061.
30. Ponce-Bobadilla AV, Schmitt V, Maier C, Mensing S. Practical guide to SHAP analysis: explaining supervised machine learning model predictions in drug development. *Clin Transl Sci*. 2024;17(11):e70056.
31. Pedregosa F, Varoquaux G, Gramfort V, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. *J Mach Learn Res*. 2011;12:2825–30.
32. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5–32.
33. Dietterich TG. Ensemble methods in machine learning. In: *International Workshop on Multiple Classifier Systems*; 2000 Jun; Berlin, Heidelberg. p. 1–15.
34. Ganaie MA, Hu M, Malik AK, Tanveer M, Suganthan PN. Ensemble deep learning: a review. *Eng Appl Artif Intell*. 2022;115:105151.
35. Sagi O, Rokach L. Ensemble learning: a survey. *Wiley Interdiscip Rev Data Min Knowl Discov*. 2018;8(4):e1249.
36. Du J, Chang X, Ye C, Zeng Y, Yang S, Wu S, Li L. Developing a hypertension visualization risk prediction system utilizing machine learning and health check-up data. *Sci Rep*. 2023;13:18953.
37. Bisong E, Jibril N, Premnath P, Buligwa E, Oboh G, Chukwuma A. Predicting high blood pressure using machine learning models in low- and middle-income countries. *BMC Med Inform Decis Mak*. 2024;24:234.
38. Islam H, Iqbal MS, Hossain MM. Blood pressure abnormality detection and interpretation utilizing explainable artificial intelligence. *Intell Med*. 2025;5(1):54–65.