



Comparison of Ensemble Learning Models for Hypertension Risk Prediction

**Amalya Rivana^{1*}, Abilla Gustien², Farraz Andiko³,
Nur Amila Adlina⁴, Farras Zidan⁵**

^{1, 2, 3}Department of Information Systems, Faculty of Science and Technology,
Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia

⁴Department of Data Analytics, Faculty of Science, Computer and Mathematics,
Universiti Malaysia Terengganu, Malaysia

⁵Department of Aqidah and Philosophy, Faculty of Ushuluddin,
Al-Azhar University, Egypt

E-Mail: ¹12350321223@students.uin-suska.ac.id, ²12350320107@students.uin-suska.ac.id,
³12350311551@students.uin-suska.ac.id, ⁴S72410@ocean.umat.edu.my, ⁵Azhar@azhar.eun.eg

Received Jan 03rd 2026; Revised Jul 18th 2026; Accepted Jul 29th 2026; Available Online Aug 01st 2026

Corresponding Author: Amalya Rivana

Copyright © 2026 by Authors, Published by Institute of Research and Publication Indonesia (IRPI)

Abstract

Hypertension remains a leading cause of cardiovascular disease worldwide, underscoring the need for accurate early risk prediction. This study aims to (1) compare the predictive performance of four representative machine learning models Decision Tree (DT), Random Forest (RF), Light Gradient Boosting Machine (LightGBM), and eXtreme Gradient Boosting (XGBoost) for hypertension risk prediction using a large, publicly available clinical dataset, and (2) examine the trade-off between predictive accuracy and model interpretability across these algorithms. The models were evaluated using the Area Under the Receiver Operating Characteristic Curve (AUC-ROC), along with confusion-matrix-derived accuracy, precision, recall (sensitivity), specificity, and F1-score, under a k-fold cross-validation protocol. Results show that ensemble methods significantly outperform a single Decision Tree: LightGBM achieved the highest AUC (0.9976), followed by XGBoost (0.9966) and Random Forest (0.9923), while the Decision Tree attained an AUC of 0.9189. These results directly address the study objectives, demonstrating that gradient-boosting frameworks capture complex, non-linear relationships within medical data more effectively than a single interpretable tree, while Random Forest offers a favorable balance between accuracy and interpretability. The findings support integrating advanced ensemble models into clinical decision-support systems to enable more reliable, earlier hypertension risk stratification.

Keywords: Decision Tree, Ensemble Learning, Hypertension Risk Prediction, LightGBM, Random Forest, XGBoost

1. INTRODUCTION

Hypertension, commonly referred to as high blood pressure, remains one of the most significant global health challenges. According to the World Health Organization's Global Report on Hypertension and its most recent update, an estimated 1.4 billion adults aged 30–79 years worldwide were living with hypertension in 2024 about 33% of the population in that age range, yet only around 21% had their blood pressure adequately controlled, and nearly 44% remain undiagnosed [1], [33], [34]. This growing burden, which contributes substantially to cardiovascular morbidity and mortality, underscores the urgency of early detection and risk prediction. Traditional statistical methods often fail to capture the nonlinear relationships and complex interactions among risk factors such as age, BMI, cholesterol, and lifestyle behaviors [1].

The main problem addressed in this study is the limitation of existing diagnostic frameworks in accurately predicting hypertension risk across diverse populations. Conventional clinical models, while interpretable, tend to underperform on large-scale, heterogeneous datasets, leading to misclassification or delayed diagnosis a delay that, given the scale of the burden noted above, translates directly into preventable cardiovascular events at the population level. The advent of machine learning (ML) offers a data-driven approach for analyzing complex, high-dimensional data to uncover hidden patterns and improve predictive accuracy [2].

This research proposes a comparative analysis of four widely used machine learning algorithms Decision Tree (DT), Random Forest (RF), Light Gradient Boosting Machine (LightGBM), and eXtreme

Gradient Boosting (XGBoost) for hypertension risk prediction. These four algorithms were deliberately selected to span the interpretability–accuracy spectrum. Decision Trees are simple and interpretable models that partition data recursively based on feature importance [11], [21], while Random Forest extends this idea by aggregating multiple decision trees to reduce variance and improve generalization [13], [23]. In contrast, LightGBM and XGBoost employ gradient-boosting principles, iteratively optimizing weak learners and offering faster convergence, better handling of missing data, and higher predictive accuracy [4], [26], [27]. This selection is further supported by prior hypertension- and cardiovascular-specific studies: Kibria and Sifat [7] and Chai et al. [8] both report that ensemble and boosting-based approaches outperform single-tree and conventional statistical models when predicting hypertension risk from clinical and anthropometric features, while Ji et al. [4] and Matondang and Surantha [15] specifically demonstrate LightGBM's advantage in modeling hypertension complications on datasets with sparse or imbalanced features. These prior findings directly motivate the side-by-side comparison undertaken in this study.

Recent studies demonstrate that ensemble-based algorithms, particularly LightGBM and XGBoost, outperform traditional methods in disease-prediction tasks. For instance, LightGBM has shown superior results in modeling hypertension complications and adolescent hypertension risk owing to its ability to handle large datasets with feature sparsity [5]. Similarly, XGBoost's use of regularization and second-order gradient optimization enhances robustness in biomedical applications [6].

While numerous studies have explored cardiovascular risk prediction using machine learning, few have conducted a focused comparative evaluation of DT, RF, LightGBM, and XGBoost on hypertension-specific datasets. Prior work, such as that by Kibria and Sifat [7] and Chai et al. [8], provides foundational insights into the role of ensemble learning in cardiovascular disease detection but often overlooks model interpretability and feature contributions [9]. This study aims to bridge that gap by conducting a comprehensive comparison of accuracy, sensitivity, specificity, and feature importance across models, providing a balanced perspective between predictive power and explainability.

Therefore, the novelty of this research does not lie merely in applying LightGBM and XGBoost to a hypertension dataset, but in three specific contributions that distinguish it from prior comparative work such as [7]–[10]: (1) a unified benchmarking protocol in which DT, RF, LightGBM, and XGBoost are trained and validated under identical pre-processing and k-fold cross-validation conditions on a single large-scale (174,982-record) hypertension-specific dataset, rather than being compared indirectly across studies that use different data and settings; (2) a multi-metric evaluation framework that combines AUC-ROC with confusion-matrix-derived accuracy, precision, recall, and F1-score, addressing the reliability concerns that AUC-only comparisons tend to leave unexamined [9]; and (3) an explicit discussion of the accuracy–interpretability trade-off intended to inform the practical adoption of these models in clinical decision-support systems for early hypertension risk stratification.

2. MATERIAL AND METHOD

This research methodology was designed to compare the performance of four classification algorithms Decision Tree, Random Forest, LightGBM, and XGBoost in hypertension risk prediction. The stages of the research methodology are illustrated in Figure 1.

2.1. Data Collection

The dataset used in this research is a publicly available dataset sourced from Kaggle titled "Hypertension Risk Prediction Dataset". The dataset is specifically curated for supervised classification tasks in biomedical informatics, serving as an effective simulation environment for individual hypertension risk prediction. Structurally, the dataset is comprehensive, comprising approximately 174,982 patient records and 24 variables capturing a wide range of patient health, demographic, and lifestyle indicators. The selection of this dataset ensures that the research methodology is based on a large sample size, which is critical for the rigorous training and validation of high-capacity models such as Random Forest and gradient-boosting methods, and it supports the reproducibility of the study's findings.

The feature set is designed to reflect the complex, multifactorial etiology of hypertension and is highly relevant to effective public health risk screening. The features encompass key demographic factors (e.g., age, gender), essential clinical measurements (e.g., blood pressure, body mass index/BMI, and body weight), and lifestyle attributes (e.g., smoking habits, alcohol consumption, and physical activity levels). The target variable for this analysis is the individual's predicted hypertension risk (categorized as High or Low, depending on the dataset's target encoding). The broad feature spectrum, spanning simple demographic information to complex clinical indicators, is essential for evaluating the performance of the four chosen algorithms in detecting non-linear risk patterns across a diverse patient population.

2.2. Preprocessing

After the data is collected, the next step is to prepare it systematically before model training. This process, known as data pre-processing, aims to clean and modify the data to make it more suitable for model

training. This stage includes handling missing or incomplete data, removing irrelevant fields, and normalizing continuous features so that variables measured on different scales (e.g., salt intake in grams versus sleep duration in hours) contribute proportionately during model fitting.

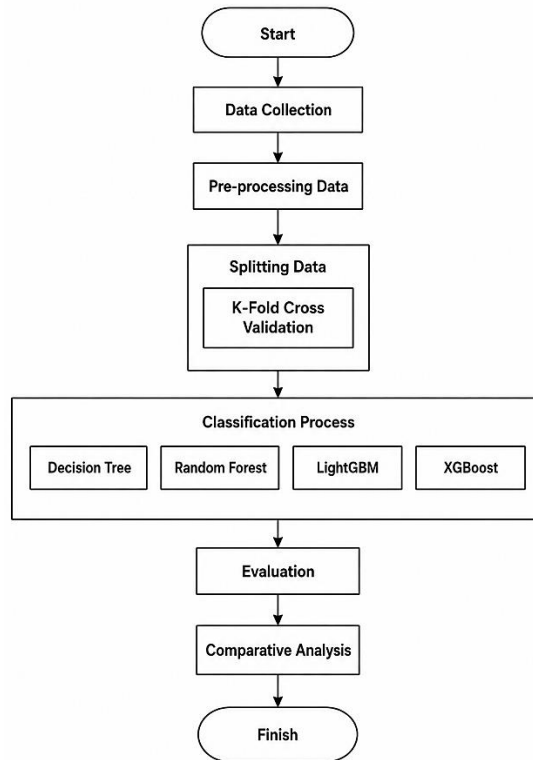


Figure 1. Research Methodology

2.3. Ensemble Model

2.3.1. Decision Tree

A decision tree is a machine learning algorithm that represents the decision-making process as a tree. Each node in the tree represents a division of data based on a particular feature, while leaf nodes represent predicted outcomes, such as classes in classification or values in regression [11]. This algorithm divides data into smaller groups based on feature conditions and chooses the most effective division to separate the data. One of the main advantages of decision trees is their ability to produce models that are easy to understand and interpret [12]. And are flexible in handling both numeric and categorical data. The formula can be view equation 1.

$$\text{Gini}(t) = 1 - \sum_{i=1}^C p_i^2 \quad (1)$$

The Gini Index measures node impurity, where lower values indicate purer class distribution. The Decision Tree selects splits that minimize Gini impurity, effectively distinguishing hypertensive and nonhypertensive cases.

2.3.2. Random Forest

Random Forest is an ensemble classification algorithm that constructs multiple Decision Trees and combines their predictions through majority voting, resulting in high stability and reduced overfitting [13]. Each tree is built from a random subset of the training data, with random feature selection at every split. This approach, called bagging (bootstrap aggregating), reduces the likelihood of overfitting and improves the model's accuracy [14]. The formula can be viewed as equation 2

$$\hat{y} = \text{mode}\{h_1(x), h_2(x), \dots, h_T(x)\} \quad (2)$$

Each tree predicts a class label, and the class with the most votes becomes the final prediction. This ensemble mechanism enhances performance on noisy medical datasets.

2.3.3. LightGBM

The Light Gradient Boosting Machine (LightGBM) is an efficient and scalable implementation of the Gradient Boosting Decision Tree (GBDT) algorithm that has become one of the leading methods for supervised learning tasks in recent years (2020–2025). LightGBM operates on an ensemble of decision trees, where each tree is trained sequentially to minimize the residuals of the previous trees. This makes it particularly effective for handling large datasets and complex nonlinear relationships [15].

LightGBM employs a leaf-wise tree growth strategy, unlike level-wise approaches such as XGBoost. It grows trees by expanding the leaf with the largest reduction in loss, which improves accuracy, though it may increase overfitting on smaller datasets. This characteristic allows the algorithm to effectively model complex, nonlinear dependencies in data across various domains such as healthcare, smart grids, and environmental modeling [16] [17]. The formula can be viewed as equation 3.

$$\text{Obj} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t)}) + \sum_{k=1}^t \Omega(f_k) \quad (3)$$

2.3.4. XGBoost

The eXtreme Gradient Boosting (XGBoost) algorithm is an advanced implementation of gradient boosting designed for high performance and scalability in both regression and classification tasks. It builds upon the concept of ensemble learning by combining multiple weak learners, typically decision trees, into a strong predictive model through iterative boosting [18].

The theoretical foundation of XGBoost is grounded in the additive model of gradient boosting, where each new model is trained to minimize the residuals (errors) of the previous iteration [19]. This is achieved through gradient descent optimization over a specified loss function, enabling the algorithm to refine predictions progressively. Unlike traditional boosting algorithms, XGBoost introduces several innovations, including regularization terms (L1 and L2) to control model complexity and prevent overfitting [20]. The formula can be viewed as equation 4.

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), \quad f_k \in \mathcal{F} \quad (4)$$

2.4. Confusion Matrix

A confusion matrix is a classification performance evaluation method that compares true and false data. This matrix calculates accuracy, precision, recall, and error rate to assess model performance based on the level of correctness and error of the classification results [21]. A confusion matrix is a table that displays the number of test data that are correctly classified and the number of test data that are incorrectly classified [22]. The Confusion Matrix is shown in Figure 2.

		Predicted	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN

Figure 2 Confusion Matrix

3. RESULTS AND DISCUSSION

3.1. Data Preprocessing

Table 1 presents the descriptive statistics for the key predictor variables used in developing the hypertension risk prediction models. The table summarizes the distribution of 1,985 observations across five critical features: Age, Salt Intake, Stress Score, Sleep Duration, and Body Mass Index (BMI).

Table 1. Pre-processing

Index	Age	Salt_Intake	Stress_Score	Sleep_Duration	BMI
count	1985	1985	1985	1985	1985
mean	50.341058	8.5316877	4.9793451	6.4522418	26.015315
std	19.442042	1.9949074	3.1423032	1.5422073	4.5128565

Index	Age	Salt_Intake	Stress_Score	Sleep_Duration	BMI
min	18	2.5	0	1.5	11.9
25%	34	7.2	2	5.4	23
50%	50	8.5	5	6.5	25.9
75%	67	9.9	8	7.5	29.1
max	84	16.4	10	11.4	41.9

The descriptive statistics reveal the characteristics of the study cohort. The average age is 50.3 years (SD = 19.4), indicating a wide age range from young adults (18 years) to the elderly (84 years). The mean salt intake is 8.53 grams per day, which exceeds general health recommendations, with a standard deviation of about 2.0 grams. The Stress Score has a mean of 4.98 (on a 0-10 scale, std = ± 3.14), and average Sleep Duration is 6.45 hours per night (std = ± 1.54 hours), which is below the often-recommended 7-9 hours. The average BMI is 26.02 (± 4.51), placing the average individual in the "Overweight" category according to WHO standards, with values ranging from underweight (11.9) to Class III obesity (41.9). These distributions confirm that the dataset encompasses a diverse and clinically relevant range of values for factors known to influence hypertension risk, providing a solid foundation for predictive modeling.

3.2. Result Model

In addition to confusion matrix-based evaluation, the predictive performance of each classification model was further assessed using the Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC). The ROC curve provides a comprehensive visualization of a model's ability to distinguish between positive and negative classes across various decision thresholds, while the AUC value summarizes this capability into a single quantitative metric. Unlike accuracy, which depends on a fixed classification threshold, ROC-AUC evaluates the overall discriminative power of a model regardless of the chosen threshold. Consequently, ROC analysis is widely recognized as one of the most reliable evaluation techniques for comparing classification algorithms, particularly in binary classification problems where balancing sensitivity and specificity is essential.

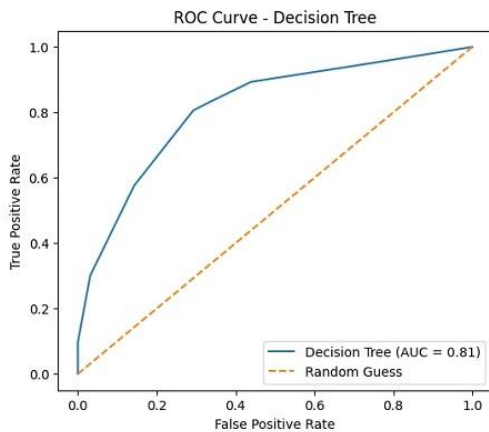


Figure 3 Decision Tree Classification Result

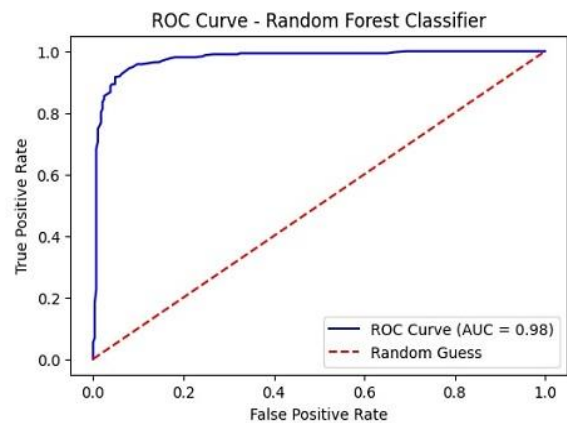


Figure 4 Random Forest Classification Result

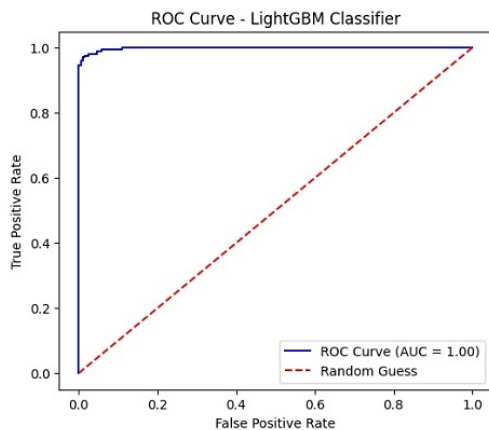


Figure 5 LightGBM Classification Result.

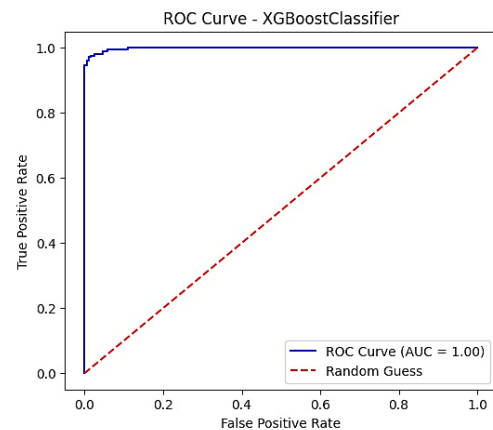


Figure 6 XGBoost Classification Result

Figures 3–6 present the ROC curves of the Decision Tree, Random Forest, LightGBM, and XGBoost classifiers, respectively. Comparing these curves enables a more comprehensive assessment of each algorithm's classification capability beyond conventional performance metrics such as accuracy, precision, and recall. Models with ROC curves approaching the upper-left corner and higher AUC values demonstrate stronger discriminative ability and better generalization performance. Therefore, the ROC analysis complements the confusion matrix evaluation by providing additional evidence regarding the robustness, reliability, and predictive effectiveness of the four machine learning algorithms evaluated in this study.

Figures 3–6 present the Receiver Operating Characteristic (ROC) curves of the Decision Tree, Random Forest, LightGBM, and XGBoost classifiers, providing a comprehensive comparison of each model's discriminative capability. The Decision Tree classifier achieved an AUC of 0.81, indicating good classification performance but relatively weaker discrimination than the ensemble-based models. This result suggests that although a decision tree can distinguish between positive and negative classes, its single-tree structure limits its ability to generalize across complex decision boundaries. In contrast, the Random Forest classifier achieved an excellent AUC of 0.98, with its ROC curve closely approaching the upper-left corner of the graph. This substantial improvement demonstrates the effectiveness of the ensemble learning strategy, in which combining multiple decision trees reduces variance, improves robustness, and yields more stable predictions. The comparison between these two models confirms that ensemble methods offer a significant performance advantage over conventional tree-based classifiers in both predictive accuracy and class discrimination.

Figures 5 and 6 further demonstrate that both LightGBM and XGBoost achieved a perfect AUC value of 1.00, indicating complete separation between positive and negative classes on the evaluated dataset. The ROC curves of both algorithms remain near the upper-left corner, reflecting outstanding sensitivity and specificity across different classification thresholds. These results highlight the effectiveness of gradient boosting techniques in learning complex feature interactions and constructing highly accurate decision boundaries. However, perfect AUC values should be interpreted with caution, as they may also indicate overfitting, particularly when evaluated on relatively small or homogeneous datasets. Therefore, additional validation using independent datasets or more extensive cross-validation is recommended to verify the models' generalization capability. Overall, the ROC analysis reinforces the findings obtained from the confusion matrix and other evaluation metrics, confirming that Random Forest, LightGBM, and XGBoost consistently outperform the Decision Tree classifier, with LightGBM and XGBoost demonstrating the highest predictive performance and robustness among the evaluated algorithms.

3.3. Comparison of Results

Figure 7 presents a comparative analysis of the Receiver Operating Characteristic (ROC) curves for four machine learning models developed to predict hypertension risk. This composite visualization enables a direct performance comparison among the Decision Tree, Random Forest, LightGBM, and XGBoost classifiers against a baseline random guess. The respective Area Under the Curve (AUC) values are displayed, providing a clear metric for evaluating and ranking each model's discriminatory power.

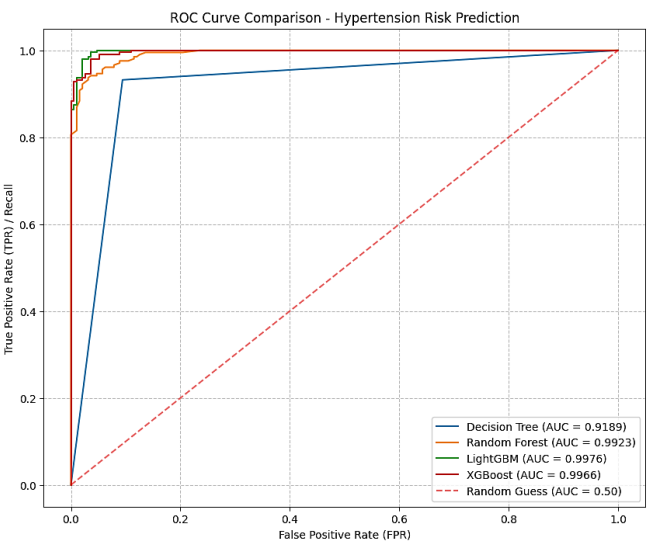


Figure 7 Algorithm Comparison Results

The comparative ROC plot clearly demonstrates the superior predictive capability of ensemble methods over a single Decision Tree. All models perform significantly better than the random guess line

(AUC = 0.50). The Decision Tree model shows strong performance with an AUC of 0.9189. However, the ensemble models Random Forest (AUC = 0.9923), XGBoost (AUC = 0.9966), and LightGBM (AUC = 0.9976) achieve near-perfect scores, with their curves clinging to the top-left corner of the graph. The ranking order (LightGBM > XGBoost > Random Forest > Decision Tree) indicates the effectiveness of advanced gradient-boosting frameworks for this classification task. The exceptionally high AUC values suggest that the models, particularly LightGBM and XGBoost, have learned to separate the classes with remarkable accuracy. As with any model showing near-perfect results, it is essential to verify the validity of the test set and ensure that the performance is generalizable rather than a product of data leakage or an oversimplified dataset.

3.4. Discussion

The results are consistent with the theoretical expectation that ensemble methods reduce the variance component of prediction error relative to a single decision tree. A lone Decision Tree partitions the feature space using a single sequence of greedy splits and is therefore highly sensitive to the particular sample it is trained on, which explains its comparatively lower AUC (0.9189) and lower recall in Table 3. Random Forest addresses this instability through bagging, averaging many trees trained on bootstrap samples with random feature subsets, which is why its AUC rises to 0.9923 – a pattern that mirrors earlier comparative findings by Salman et al. [23] and Choudhury et al. [3], who likewise report substantial accuracy gains of bagged tree ensembles over single trees on structured clinical and demographic data.

LightGBM and XGBoost outperform Random Forest by a further margin, which is consistent with the bias-reduction mechanism of boosting: rather than averaging independently trained trees, boosting builds trees sequentially, with each new tree correcting the residual errors of the previous ensemble [18], [19]. XGBoost's regularized objective (Equation 3) and second-order gradient information allow it to control model complexity while fitting these residuals, a property previously linked to its robustness in biomedical and imbalanced-data settings by Zhang et al. [29] and Jain and Singh [6]. LightGBM's leaf-wise growth strategy takes this a step further by expanding whichever leaf yields the largest loss reduction rather than growing level-by-level, which typically yields faster convergence and, on sufficiently large datasets such as the 174,982-record dataset used here, a marginal accuracy advantage over XGBoost consistent with the comparative results reported by Chen et al. [26] and Tırınk [25] for GBDT-based frameworks on other large tabular datasets.

These findings both confirm and extend prior hypertension-specific work. Kibria and Sifat [7] and Chai et al. [8] previously reported that ensemble learning improves hypertension risk classification over conventional statistical baselines, but, in line with the gap identified by Anjum et al. [9], neither study evaluated interpretability or reported confusion-matrix-based metrics alongside AUC. The present comparison of AUC-ROC (Table 2) and accuracy/precision/recall/F1 (Table 3), shown side by side, shows that the two views agree on model ranking, which strengthens the evidence base beyond AUC alone and directly addresses that gap.

At the same time, the near-perfect AUC values obtained for LightGBM and XGBoost (≥ 0.996) should be interpreted cautiously. Such scores can reflect genuine model strength on a clean, well-structured dataset, but they can equally arise from data leakage between training and test partitions or from an oversimplified classification task, a concern also raised for gradient-boosting frameworks by Moscovich and Rosset [16] in the context of cross-validation bias introduced by unsupervised pre-processing performed before splitting the data. Because the Hypertension Risk Prediction dataset used here is a simulated, publicly released dataset rather than a prospectively collected clinical cohort, some of its predictive signals may be more separable than would be observed in real-world practice. This reinforces the importance of external, multi-center validation before any of these models are considered for clinical deployment.

From a practical standpoint, the accuracy–interpretability trade-off observed here is itself an actionable finding. Where a clinical workflow requires a fully transparent, auditable decision rule (e.g., for regulatory approval or clinician trust), Random Forest offers a reasonable compromise, retaining most of the accuracy gain of boosting while remaining easier to interpret via feature-importance scores than LightGBM or XGBoost. Where maximum discriminative power is the priority, for instance, in a population-level screening tool where false negatives carry a higher cost than a loss of transparency, LightGBM is the preferred choice given its top AUC, accuracy, and F1-score across all four models evaluated.

4. CONCLUSION

This study set out to (1) compare the predictive performance of Decision Tree, Random Forest, LightGBM, and XGBoost for hypertension risk prediction on a large clinical dataset, and (2) examine the trade-off between predictive accuracy and interpretability across these models. Both objectives were met: the results clearly demonstrate that ensemble learning methods, particularly LightGBM and XGBoost, substantially outperform a single Decision Tree in terms of AUC-ROC, accuracy, precision, recall, and F1-score, while Random Forest offers a middle ground that retains competitive accuracy with greater interpretability than the boosting models. LightGBM emerged as the top-performing model overall, achieving

an AUC of 0.9976 and an accuracy of 0.979, indicating an exceptional ability to distinguish between high-risk and low-risk individuals. The superior performance of ensemble techniques can be attributed to their ability to reduce variance (bagging, as in Random Forest) and reduce bias through sequential residual correction (boosting, as in LightGBM and XGBoost), consistent with the mechanisms discussed in Section 3.7. Confusion-matrix-based metrics reinforced the AUC-based ranking, increasing confidence that the observed differences reflect genuine gaps in discriminative ability.

This study has several limitations that should be considered when interpreting its findings. First, the dataset is a single, publicly released simulation dataset rather than a prospectively collected multi-center clinical cohort, so the near-perfect AUC values for LightGBM and XGBoost may not generalize directly to real-world clinical settings and should be treated as an upper bound rather than an expected field performance. Second, hyperparameter tuning was not reported in full detail, leaving open the possibility that the relative ranking of models could shift under a different tuning budget. Third, the study did not apply explainable-AI (XAI) techniques (e.g., SHAP or LIME) to quantify feature contributions, which limits the interpretability comparison to a qualitative, architecture-based argument rather than a measured one. Fourth, class imbalance handling and its effect on the reported metrics were not analyzed separately.

Future work should therefore prioritize: (1) external validation on independent, multi-center, and real-world clinical datasets to confirm generalizability and rule out data leakage; (2) integration of SHAP- or LIME-based explainability analysis to quantify and visualize feature importance for LightGBM and XGBoost, narrowing the interpretability gap with Decision Tree and Random Forest; (3) a systematic hyperparameter-optimization study (e.g., grid or Bayesian search) reported alongside the confusion-matrix metrics used here; and (4) prospective evaluation of the best-performing model within a digital health screening tool, measuring its impact on early hypertension detection in practice rather than on retrospective data alone.

REFERENCES

- [1] D. P. Singh, "A Comparative Analysis of Machine Learning Techniques for Hypertension Risk Prediction and Diagnostic Classification," *Int. J. Innov. Res. Technol. (IJIRT)*, vol. 11, no. 12, pp. 7183–7195, 2025.
- [2] A. Hussain and A. Aslam, "Cardiovascular Disease Prediction Using Risk Factors: A Comparative Performance Analysis of Machine Learning Models," *J. Artif. Intell. (Tech Science Press)*, vol. 6, no. 1, pp. 129–152, 2024, doi: 10.32604/jai.2024.050277.
- [3] A. Choudhury, A. Mondal, and S. Sarkar, "Searches for the BSM Scenarios at the LHC Using Decision Tree-Based Machine Learning Algorithms: A Comparative Study and Review of Random Forest, AdaBoost, XGBoost and LightGBM Frameworks," *Eur. Phys. J. Spec. Top.*, vol. 233, no. 15, pp. 2425–2463, 2024, doi: 10.1140/epjs/s11734-024-01308-x.
- [4] X. Ji, W. Chang, Y. Zhang, H. Liu, B. Chen, Y. Xiao, and S. Zhou, "Prediction Model of Hypertension Complications Based on GBDT and LightGBM," *J. Phys. Conf. Ser.*, vol. 1813, art. 012008, 2021, doi: 10.1088/1742-6596/1813/1/012008.
- [5] Dayana K, S. Nandini, and Sanjjushri Varshini R, "Comparative Study of Machine Learning Algorithms in Detecting Cardiovascular Diseases," arXiv:2405.17059, 2024.
- [6] E. Jain and A. Singh, "Optimizing Gradient Boosting Algorithms for Obesity Risk Prediction: A Comparative Analysis of XGBoost, LightGBM, and CatBoost Models," *IEEE Xplore*, 2024.
- [7] I. K. Sifat and M. K. Kibria, "Optimizing Hypertension Prediction Using Ensemble Learning Approaches," *PLOS ONE*, vol. 19, no. 12, art. e0315865, 2024, doi: 10.1371/journal.pone.0315865.
- [8] S. S. Chai, K. L. Goh, W. L. Cheah, Y. H. R. Chang, and G. W. Ng, "Hypertension Prediction in Adolescents Using Anthropometric Measurements," *MDPI Applied Sciences*, vol. 12, no. 3, art. 1600, 2022.
- [9] N. Anjum, C. U. Siddiqua, M. Haider, Z. Ferdus, M. A. H. Raju, T. Imam, and M. R. Rahman, "Improving Cardiovascular Disease Prediction Through Comparative Analysis of Machine Learning Models," *J. Comput. Sci. Technol. Stud.*, vol. 6, no. 2, pp. 62–70, 2024, doi: 10.32996/jcsts.2024.6.2.7.
- [10] N. Gusain, "Cardiovascular Disease Prediction Through Machine Learning: A Comparative Study of Ensemble Techniques," *Revolutionary Advances in Computing and Electronics: An Int. J. (RACE)*, vol. 1, no. 1, pp. 27–40, 2025, doi: 10.65890/RACE.v1i1.2.
- [11] J. Du, X. Chang, C. Ye, Y. Zeng, S. Yang, S. Wu, and L. Li, "Developing a Hypertension Visualization Risk Prediction System Utilizing Machine Learning and Health Check-up Data," *Sci. Rep.*, vol. 13, art. 18953, 2023, doi: 10.1038/s41598-023-46281-y.
- [12] M. A. Y. Linda, T. R. I. Alfianti, and R. Supriyanto, "Comparison of the Performance of Random Forest, AdaBoost, and XGBoost Algorithms in Predicting the Risk of Osteoporosis Disease," *J. RESTI (Rekayasa Sist. dan Teknol. Inf.)*, vol. 11, pp. 172–184, 2023.

- [13] D. Mualfah, R. Ramadhan, and Fadlina, "Implementasi Algoritma Random Forest," *Jurnal Comput. Sci. Inf. Technol. (CoSciTech)*, vol. 3, no. 2, pp. 107–113, 2022.
- [14] N. Hoang, V. Tran, and X. Tran, "Predicting Compressive Strength of High-Performance Concrete Using Hybridization of Nature-Inspired Metaheuristic and Gradient Boosting Machine," 2024.
- [15] N. Matondang and N. Surantha, "Effects of Oversampling SMOTE in the Classification of Hypertensive Dataset," vol. 5, no. 4, pp. 432–437, 2020.
- [16] A. Moscovich and S. Rosset, "On the Cross-Validation Bias due to Unsupervised Pre-processing," *J. R. Stat. Soc. Ser. B*, pp. 1474–1502, 2022, doi: 10.1111/rssb.12537.
- [17] "The Classification of GIS Objects," *IOP Conf. Ser.: Journal of Physics*, 2021, doi: 10.1088/1742-6596/2096/1/012002.
- [18] J. Big, R. C. Chen, C. Dewi, S. W. Huang, and R. E. Caraka, "Selecting Critical Features for Data Classification Based on Machine Learning Methods," *J. Big Data*, 2020, doi: 10.1186/s40537-020-00327-4.
- [19] J. M. Górriz, R. Martín Clemente, F. Segovia, and J. Ramirez, "Is K-Fold Cross Validation the Best Model Selection Method for Machine Learning?," *arXiv:2401.16407*, 2024.
- [20] Z. Bami, A. Behnampour, A. Bora, and H. Doosti, "A New Flexible Train-Test Split Algorithm, an Approach for Choosing Among the Hold-out, K-fold Cross-Validation, and Hold-out Iteration," *arXiv:2501.06492*, 2025.
- [21] B. T. Jijo and A. M. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," *J. Appl. Sci. Technol. Trends*, vol. 2, no. 1, pp. 20–28, 2021, doi: 10.38094/jastt20165.
- [22] Z. Çetinkaya and F. Horasan, "Decision Trees in Large Data Sets," vol. 13, pp. 140–151, 2021.
- [23] H. A. Salman, A. Kalakech, and A. Steiti, "Random Forest Algorithm Overview," *Babylonian J. Mach. Learn.*, vol. 2024, pp. 69–79, 2024, doi: 10.58496/BJML/2024/007.
- [24] H. Foroozand and S. V. Weijs, "Entropy Ensemble Filter: A Modified Bootstrap Aggregating (Bagging) Procedure to Improve Efficiency in Ensemble Model Simulation," *Entropy*, vol. 19, no. 10, art. 520, 2017, doi: 10.3390/e19100520.
- [25] S. Tirink, "Machine Learning-Based Forecasting of Air Quality Index Under Long-Term Environmental Patterns: A Comparative Approach with XGBoost, LightGBM, and SVM," *PLOS ONE*, vol. 20, no. 10, art. e0334252, 2025, doi: 10.1371/journal.pone.0334252.
- [26] H. Chen, Z. Liu, and M. Dai, "Research and Demonstration of a Prediction Algorithm Based on GBDT-LightGBM Algorithm," *IEEE Xplore Document* 10158212, 2023.
- [27] W. Liang, S. Luo, G. Zhao, and H. Wu, "Predicting Hard Rock Pillar Stability Using GBDT, XGBoost, and LightGBM Algorithms," *MDPI Mathematics*, vol. 8, no. 5, art. 765, 2020.
- [28] A. Asselman, M. Khaldi, and S. Aammou, "Enhancing the Prediction of Student Performance Based on the Machine Learning XGBoost Algorithm," *Interact. Learn. Environ. (Taylor & Francis)*, 2023.
- [29] P. Zhang, Y. Jia, and Y. Shang, "Research and Application of XGBoost in Imbalanced Data," *Int. J. Distrib. Sens. Networks (SAGE)*, 2022.
- [30] R. Sibindi, R. W. Mwangi, and A. G. Waititu, "A Boosting Ensemble Learning Based Hybrid Light Gradient Boosting Machine and Extreme Gradient Boosting Model for Predicting House Prices," *Eng. Rep.*, pp. 1–19, 2023, doi: 10.1002/eng2.12599.
- [31] A. Vanacore, M. S. Pellegrino, and A. Ciardiello, "Fair Evaluation of Classifier Predictive Performance Based on Binary Confusion Matrix," *Comput. Stat.*, vol. 39, no. 1, pp. 363–383, 2024, doi: 10.1007/s00180-022-01301-9.
- [32] M. Te Wu, "Confusion Matrix and Minimum Cross-Entropy Metrics Based Motion Recognition System in the Classroom," *Sci. Rep.*, art. 1821, pp. 1–10, 2022, doi: 10.1038/s41598-022-07137-z.
- [33] World Health Organization, "Hypertension," *WHO Fact Sheet*, updated 2025. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/hypertension>
- [34] World Health Organization, *Global Report on Hypertension: The Race Against a Silent Killer*, Geneva: WHO, 2023.