https://intjhs.org/

International Journal of Health Sciences

Original Article

Advanced machine learning algorithms for blood pressure classification: Early detection or prevention could save lives

Jonathan Iworiso, PhD[1]*, Bari-ika Nornubari Vite, PhD[2], Itunu Godwin Osuntoki, PhD[3], Idorenyin A. Amaunam, MSc[4], Idris Olayiwola Olawale, MSc[1], Saksham Arora, MSc[1]

[1]School of Computing and Digital Media, London Metropolitan University, London, [2]Department of Applied Social Sciences/Health and Social Care, University of Bedfordshire, Milton Keynes, [3]Department of Statistics, Modelling and Economics, UK Health Security Agency, London, [4]School of Computer Science, University of Essex, Colchester, United Kingdom.

***Corresponding author:**
Jonathan Iworiso,
School of Computing and Digital Media, London Metropolitan University,
London, United Kingdom.

j.iworiso@londonmet.ac.uk

Received: 16 November 2024
Accepted: 28 April 2025
Published: 01 July 2025

DOI
10.25259/OJS_8819

Quick Response Code:

ABSTRACT

Objectives: The primary objective of the study is to classify the blood pressure (BP) levels using advanced machine learning (ML) techniques for predictive purposes. The study assesses the efficacy of the Naïve Bayes, AdaBoost, feedforward neural networks (FNNs), and long short-term memory (LSTM) algorithms over the conventional multinomial logistic model using standard performance evaluation metrics.

Methods: The dataset comprised 15,000 entries obtained from the National Health Service, England, each containing eight variables. The variables include BP, age, weight, height, gender, smoking habit, alcohol consumption, and fitness level. The Naïve Bayes, AdaBoost, FNN, LSTM, and multinomial logistic models were employed in the study. Each model underwent training, testing, validation, and evaluation using suitable metrics such as accuracy, F1-Score, kappa statistics, sensitivity, specificity, and area under the curve score.

Results: The FNN model gives the highest test accuracy of 89.47% and balanced performance, making it the most appropriate model for predicting BP levels. The LSTM model demonstrated strong proficiency in capturing temporal patterns. AdaBoost was highly effective for dealing with class imbalance, but Naïve Bayes was a dependable benchmark. The multinomial logistic model established a reliable and stable reference point. The results represented a notable improvement over previous research, which typically reported median accuracy rates in the 80–85% range.

Conclusion: The study reveals that knowing an individual's age, weight, height, gender, smoking habit, alcohol consumption, and fitness level is useful in predicting his/her BP level. Thus, the advanced ML algorithms demonstrate potential in accurately classifying BP levels and can aid in the prevention, detection, and management of hypertension.

Keywords: AdaBoost, Blood pressure, Feedforward neural network, Long short-term memory, Multinomial, Naïve Bayes

INTRODUCTION

The categorization of blood pressure (BP) is a crucial component of clinical research. BP measurements, acquired during physical examinations, outpatient appointments, or hospital

How to cite this article: Iworiso J, Vite BN, Osuntoki IG, Amaunam IA, Olawale IO, Arora S. Advanced machine learning algorithms for blood pressure classification: Early detection or prevention could save lives. Int J Health Sci (Qassim). 2025;19:31-42. doi: 10.25259/OJS_8819

stays, form the foundation for most medical interventions. The measurements of these readings are expressed in millimeters of mercury (mmHg) and comprise two values: The systolic pressure (numerator) and the diastolic pressure (denominator).[1] A noticeable research obstacle in determining BP is the identification of characteristics that can accurately predict an individual's BP level. Previous studies have suggested that variables such as age, heredity, body mass, stature, physical activity, smoking patterns, sodium consumption, and alcohol use can influence BP levels. Elevated BP is a prominent risk factor for cardiovascular disease and cerebrovascular accident.[2] According to recent data from National Health Service (NHS) England and the Office for National Statistics, almost 30% of adults in the UK receive a diagnosis of high BP from their general practitioner, while 29% have untreated hypertension.[3] The prevalence of high BP in England is 31% among males and 26% among females. Approximately 75 million individuals in the United States have hypertension.[4,5] It is crucial to emphasize that sporadic episodes of elevated BP caused by sickness or stress do not automatically signify hypertension.

Previous studies have shown the categorization of BP into a binary or categorical form, utilizing logistic regression with only two classes.[6] The response variable is high BP or hypertension, and it is assumed to follow a binomial distribution with parameters 1 and p, where p represents a probabilistic measure. A person is typically categorized as either having high BP (0 = yes) or not having high BP (1 = no).[7,8] Nevertheless, this strategy primarily emphasizes a seemingly binary result that lacks informative or precautionary details for individuals. For example, a person's BP may be identified as elevated yet categorized as within the normal range, regardless of the potential risk of progressing to hypertension. Furthermore, the investigation into the potential association between hypertension and other categorical variables, such as smoking habit, physical exercise, and alcohol consumption, will be limited to only two categories for the hypertension variable. This limitation may not yield enough evidence to effectively protect individuals from the risk factors of high BP and related cardiovascular diseases. A recent investigation has demonstrated the effectiveness of multiclass logistic regression and traditional machine learning (ML) methods for BP categorization,[9,10] over previous studies regarding the classification and management of hypertension.[11-14] The results seem to be promising, but the predictive accuracy could be improved. Enhancements could be made by choosing more pertinent explanatory factors and utilizing more advanced algorithms that have the potential to surpass the standard methodologies employed in the investigations.

The present study reports novel multiclass advanced ML algorithms designed for the classification of BP levels. Before utilizing advanced algorithms, we establish a model for predicting multiclass BP using multinomial methods with relevant explanatory variables. The categorization of the four groups is derived from the guidelines established by the American College of Cardiology and American Heart Association.[9]

Since 2005, the World Hypertension League has been leading a global movement to raise awareness about the importance of hypertension through yearly screening programs. The 2023 survey findings indicated that of the 502,079 persons diagnosed with hypertension, only 59.5% had knowledge of their disease.[15] This indicates a requirement for enhanced consciousness and instruction around hypertension. The existing research highlights the inadequacy of the present hypertension screening programs used by the NHS. This highlights the need for improvements in the effectiveness of these initiatives. To successfully tackle this pressing matter, it is imperative to devise novel ways for enhancing the identification of hypertension on a population scale. Through the identification of risk factors for hypertension, we can proactively undertake interventions at an early stage to avoid its occurrence, recognize it promptly, and mitigate the long-term repercussions linked to it. The expertise of healthcare professionals, researchers, and policymakers is essential in this undertaking. Their contributions have the capacity to generate a substantial influence on the lives of numerous individuals.

The British and Irish Hypertension Society hosts a yearly nationwide initiative to raise awareness about the critical importance of identifying high BP. Various surveys have been utilized to gather data about hypertension throughout the years. A total of 37,110 persons actively engaged in the World Hypertension Day initiatives between 2020 and 2024, resulting in a significant dataset. After examining the initial dataset, a total of 20,206 participants were meticulously chosen to be included in the current study. This selection approach entailed the exclusion of persons who had previously been diagnosed with hypertension. Out of the chosen subjects, 4192 (20.75%) individuals were identified as having been recently diagnosed with hypertension. The dataset included demographic information, risk factors, general knowledge inquiries about hypertension, and three measurements of systolic and diastolic BP, as well as heart rate. The question is, "what about those individuals who are unlikely to present themselves for vital signs checks, medical tests, or diagnosis"? Knowing the age, weight, height, smoking habit, fitness level, and other relevant factors could be helpful in predicting the individual's hypertensive status regardless of their negligence for medical tests. An early detection approach in this sense could trigger a follow-up to enhance the individual's medical safety, thereby preventing the individual from moving from a normal or lower stage to a higher stage of hypertension.

In recent years, artificial intelligence (AI) has been successfully employed in health care as a beneficial medical instrument for numerous clinical diseases.[16,17] ML excels at generating precise predictions about individual outcomes by leveraging acquired data to identify learned patterns, rather than depending on explicit programming. Multiple ML algorithms have been employed in the statistical predictability of hypertension, resulting in varied outcomes across different research.[18-23]

Thus, the objective of this study was to employ supervised ML techniques with a substantial dataset to create algorithms capable of precisely detecting unidentified hypertensive status. In this study, we aimed to evaluate the effectiveness of these models or algorithms in comparison to the current screening techniques. The research procedure is comprehensive, utilizing five advanced ML classifiers. The models are subjected to a 10-fold cross-validation on the original training set, applying oversampling and undersampling procedures to resolve any imbalances in the data. This approach is likely to instill trust in the capability of ML in identifying hypertensive status from the multiclass. The evaluation of model performance entails a comparative analysis of the training and testing sets using prominent performance evaluation metrics such as sensitivity, specificity, accuracy, and precision, among others.

MATERIALS AND METHODS

Data pre-processing

The dataset is split into a training set and a test set. We utilized sophisticated ML techniques with a supervised learning strategy to acquire knowledge from the data. This was achieved by training the models, cross-validating, and making predictions on the test dataset. We use the assumption that the BP conforms to a multinomial distribution with four distinct classes. The categorization of the BP levels obtained from the NHS England into four classes is based on the guidelines established by the American College of Cardiology and American Heart Association. Let $I_{BP}^{class}(Y \mid X)$ represent a binary indicator function for BP. This function is defined as a multiclass variable in the form:

$$I_{BP}^{class}(Y \mid X) = \begin{cases} 1 & Normal\ (i.e., < 120/80\ mmHg) \\ 2 & Elevated\ (i.e., 120-129/<80\ mmHg) \\ 3 & Stage\,1\,hypertension\ (i.e., 130-139/80-89\ mmHg) \\ 4 & Stage\,2\,hypertension\ (i.e., \geq 140/\geq 90\ mmHg) \end{cases} \quad (1)$$

where $I_{BP}^{class}(Y \mid X)$ follows multinomial distribution with probabilistic classifier and cutoff values defined by:

$$Prob(I_{BP}^{class}) = \begin{cases} 0.51-0.99 & Normal \\ 0.31-0.50 & Elevated \\ 0.11-0.30 & Stage\,1\,hypertension \\ 0.00-0.10 & Stage\,2\,hypertension \end{cases} \quad (2)$$

The data are split into a training set comprising 70% and a test set comprising 30%. In the first step, we apply a multinomial logistic regression model to the data. Thereafter, the feasibility of the fitted model is assessed, and we perform statistical tests to determine its significance. Subsequently, we utilize sophisticated ML classifiers to analyze the training data and make predictions for the test set. For this task, the supervised ML technique is suitable. The research employs advanced ML classifiers, namely Naïve Bayes, Adaptive Boosting (AdaBoost), feedforward neural network (FNN), and long short-term memory (LSTM). The effectiveness of the proposed algorithms was evaluated using a multiclass confusion matrix and various performance evaluation metrics, including prediction accuracy, classification accuracy, kappa statistics, F1-score, specificity, sensitivity, and precision. These metrics were compared to the results obtained from conventional techniques used in previous studies. Furthermore, we performed statistical tests of independence to further examine the potential association between hypertension and each of the categorical explanatory factors.

Naïve Bayes algorithm

The Naïve Bayes algorithm is a probabilistic classifier based on Bayes' Theorem, which underpins a broad range of applications in ML. Fundamentally, Naïve Bayes assumes that the presence of a particular feature in a class is independent of the presence of any other feature.[24] This assumption, often termed "conditional independence," simplifies the computation, and while it may seem oversimplified, Naïve Bayes classifiers have proven remarkably effective in numerous practical applications, particularly in text classification tasks such as spam detection and sentiment analysis.[25]

The algorithm leverages Bayes' Theorem, which relates the conditional and marginal probabilities of stochastic events. Specifically, for a classification problem, it calculates the posterior probability of a class based on a set of predictors. The Naïve Bayes algorithm requires that attribute pairs are conditionally independent, given the class variable.[26] In a nutshell, it is summarized as:

$$P\left(c_i \middle| \vec{b}\right) = P\left(c_i\right) * \frac{P(\vec{b} \mid c_i)}{P(c_i)}$$

where, c_i is the class, $P\ (c_i)$ is the class probability, and $\vec{b}$ is the attribute variable.

This model is computationally efficient and easy to implement, contributing to its widespread use in scenarios with large datasets and high-dimensional input spaces. The effectiveness of Naïve Bayes has been demonstrated, particularly when the conditional independence assumption holds reasonably well.[27]

Multinomial logistic model

Multinomial logistic model extends a binary logistic to multiclass problems, where the dependent variable can take three or more categories. This statistical model estimates the probabilities of the different possible outcomes of a categorically distributed dependent variable, given a set of independent variables.[28]

The model operates on the principle that each category's log-odds are a linear combination of the independent variables. Mathematically, for K possible outcomes, the probability of the outcome k is modeled as:

$$P(Y=k|X=x)=\frac{\exp(\beta_{k_0}+\beta_k^T x)}{1+\sum_{j=1}^{K-1}\exp(\beta_{j_0}+\beta_j^T x)}$$

where, β_{k0} and β_k are parameters to be estimated, and x represents the vector of independent variables. The denominator ensures that the probabilities sum to one.

Multinomial logistic regression is widely used in areas such as medical research, market research, and social sciences, where the outcomes are nominal, and the independent variables are linearly related to the log-odds of the outcomes.[29] Empirical studies have validated the utility and broad applicability of the multinomial logistic regression in interpreting complex relationships between categorical data. Researchers have postulated that it supports fewer complex models to produce accurate predictions.[30]

The AdaBoost algorithm

The AdaBoost (Adaptive Boosting) algorithm is a powerful ensemble technique that combines multiple weak classifiers to form a robust classifier. The AdaBoost algorithm recursively merged linearly to form a more satisfying classifier.[31,32] The AdaBoost is one of the most influential developments in the field of ML, specifically within the domain of classification.[31]

AdaBoost works by training a sequence of classifiers, typically decision trees, on repeatedly modified versions of the data. Each successive classifier is trained on a dataset that has been altered in terms of the distribution of the training sets. Sets that were misclassified by previous classifiers are given increased weight, thereby focusing the learning of subsequent classifiers on the harder cases.

The output of the AdaBoost algorithm is a weighted sum of the classifiers' predictions. Each classifier's weight in the sum is assigned based on its accuracy, with more accurate classifiers having a higher influence on the final decision. Mathematically, the final model can be expressed as:

$$F(x)=sign\ (\sum_{t=1}^{T}\propto_t\ h_t(x))$$

where. $h_t\ (x)$ is the prediction of the *t-th* classifier, and $\propto_t$ is its weight, calculated from its error rate on the training set.

AdaBoost's effectiveness has been demonstrated to reduce both bias and variance, leading to improved prediction accuracy over individual classifiers and other ensemble methods.[33] The algorithm's simplicity and effectiveness have made it a standard benchmark in ML tasks.

FNNs

One of the methods we are considering in this study is the FNNs. The FNN is a pivotal class of artificial neural networks characterized by non-cyclical node connections. This non-cyclical design ensures a seamless flow of data from input to output, enhancing their efficiency in tasks such as speech recognition and image classification.[34] Unlike recurrent neural networks (RNNs), which utilize feedback loops, FNN are structured in distinct layers, input, hidden, and output, that facilitate parallel processing through interconnected computational units.[35] The multilayer perceptron's (MLPs) or multilayer neural network is an important special case of the FFN, whereby the links to the i^{th} layer come only from the immediately preceding layer ($i^{th}-1$).[36,37] The input layer receives data and assigns each feature to a neuron; the hidden layers conduct complex computations to abstractly transform the input; and the output layer, often employing a SoftMax function, converts the outputs into a probability distribution for classification tasks.[38] Each neuron processes a weighted sum of inputs through a non-linear activation function like sigmoid, tanh, or ReLU, with the choice of activation function critically affecting the network's efficiency and training dynamics.[35,38] The configuration of these layers, especially the number and size of the hidden layers, significantly influences the network's performance.

Figure 1 shows a simple example of an FNN with a single hidden layer ($\bar{h}_1$) and a single-out layer, where $h_{1n}=\varnothing\left(\sum_n^i x_i w_{im}\right)$; $\emptyset$ is the activation symbol and w_{im} are the corresponding weight matrices, connecting neuron i from the input to a neuron m in the hidden layer.[35] FNN utilizes forward propagation for data processing, where neurons sequentially transmit information up to the output layer, while the learning is driven by backpropagation, where the network updates its weights to minimize the error and

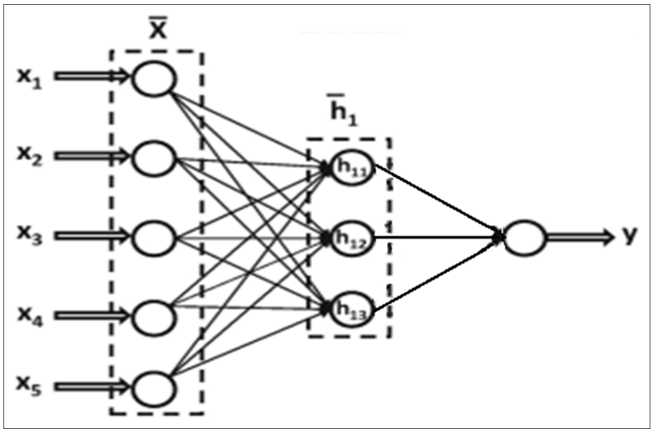

Figure 1: A simple example of a feedback network with a single hidden layer and a single output layer.

improve its predictions. Backpropagation uses a mechanism where the loss function's gradient is computed concerning each weight, guiding optimization algorithms such as SGD, Adam, or RMSprop to adjust weights and minimize errors.[36]

However, FNNs are prone to overfitting and can suffer from vanishing or exploding gradients during training, potentially leading to poor generalization and unstable updates.[39] These challenges necessitate strategic network design and the implementation of regularization techniques to bolster the robustness and efficacy of FNN in practical settings.

LSTM

LSTM networks, a specialized subclass of RNNs, were developed by Hochreiter and Schmidhuber in 1997 to tackle the challenges associated with recognizing long-term dependencies in sequential data.[40] These networks have revolutionized deep learning, significantly enhancing performance in tasks requiring comprehension of lengthy sequences, such as language translation, speech recognition, and time series analysis.[41]

The distinctive architecture of LSTM is built around a structure called the cell, which consists of three gates: The input gate, the forget gate, and the output gate. These components are crucial for regulating the flow of information. The input gate controls the entry of new data into the cell, the forget gate decides what information to retain or remove from the cell state, and the output gate determines which parts of the cell state should be output at each step.[42,43] This configuration allows LSTMs to maintain or discard data over extended periods effectively.

Figure 2 is a simple example to demonstrate the LSTM method. Where X is the pointwise multiplication, + is the pointwise addition, i_m is the input gate ($m = i$), f_i is the forget gate, o_i is the output gate, *tanh* is the pointwise Tanh, *tanh* * is the Tan h activated and σ is the sigmoid activated

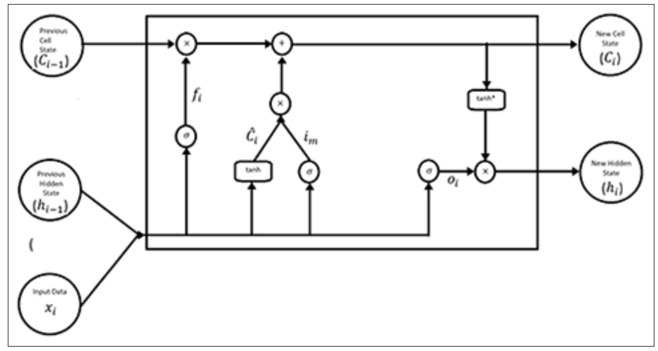

Figure 2: A simple example of the long short-term memory.

(i = 1,2,…., n)[1]. A key feature of the LSTM is the cell state, which facilitates the smooth flow of information with minimal changes.[35] This design is instrumental in mitigating the vanishing gradient problem common in traditional RNNs, where backpropagated gradients tend to decrease exponentially, hindering the learning of dependencies over longer sequences.

In practical scenarios, LSTMs excel in fields where understanding contextual continuity is crucial.[41] They outperform conventional RNNs and other models by adeptly managing data storage and retrieval, making them ideal for complex prediction tasks involving extensive historical data. Their capacity to learn sequence dependencies underscores their superiority and adaptability in the analysis of sequential data, demonstrating the significant impact of their innovative gating mechanisms on the efficacy of temporal neural models.

RESULTS

Data overview and pre-processing

The dataset used for this study consists of 15,000 observations, each containing eight variables: age, gender, weight, smoking status, alcohol consumption, fitness level, height, and BP level. The distribution of the BP levels was skewed, with most records (11,021) falling into the BP Level 1 category (normal), followed by 2,889 records in BP Level 2 (elevated), 702 in BP Level 3 (Stage 1 hypertension), and 388 in BP Level 4 (Stage 2 hypertension). Data pre-processing was performed to ensure that the dataset was free of missing values and errors. A Chi-square test was used to evaluate the relationships between categorical variables and BP levels. Significant dependencies were identified for gender and smoking status (P < 0.05), while no significant relationship was found for alcohol consumption and fitness level. The dataset was split into training set (70%) and test set (30%), with the original BP level distribution maintained. Synthetic Minority Oversampling Technique (SMOTE) was applied to the training data to address class imbalance.

Model training, testing, and evaluation

AdaBoost achieved a training accuracy of 83.43% and a test accuracy of 89.24%. It performed well on BP Levels 1 and 2, but struggled with the low-frequency categories, BP Levels 3 and 4. The F1-score was 0.70, with a kappa statistic 0.77, indicating substantial agreement. The model's sensitivity was 0.81 and specificity was high at 0.97. The area under the curve (AUC) score was 0.91, reflecting strong discrimination between BP levels.

The LSTM model showed a training accuracy of 92.8% and a test accuracy of 89.24%. While the confusion matrix indicated strong performance across all categories, it had slightly lower sensitivity and F1-scores (0.66) than AdaBoost. The kappa statistic was 0.76, with a sensitivity of 0.74 and a specificity of 0.97. The AUC score was 0.85.

The FNN achieved the highest training accuracy of 93.33% and a test accuracy of 89.47%, indicating potential overfitting. It had an F1-score of 0.65, a kappa statistic of 0.77, and a sensitivity of 0.72, with high specificity (0.97). The AUC score was 0.84, indicating strong classification performance. As shown in Figure 3, the confusion matrix indicates a balanced performance with good sensitivity and specificity across all classes.

The Naïve Bayes model reached a training accuracy of 90% and a test accuracy of 89.07%. It excelled in BP Levels 1 and 2, but its performance was weaker for Levels 3 and 4. The F1-score was 0.69, with a kappa statistic of 0.79, reflecting the highest agreement among the models. Sensitivity was 0.78, while specificity remained high at 0.95, and the AUC score was 0.88.

The multinomial logistic model achieved a training accuracy of 83.21% and a test accuracy of 89.07%, making it a reliable baseline for comparison. The F1-score was 0.68, with a kappa statistic of 0.77. Sensitivity was 0.79 and specificity was 0.97, with an AUC score of 0.89, demonstrating its competence in distinguishing between BP levels.

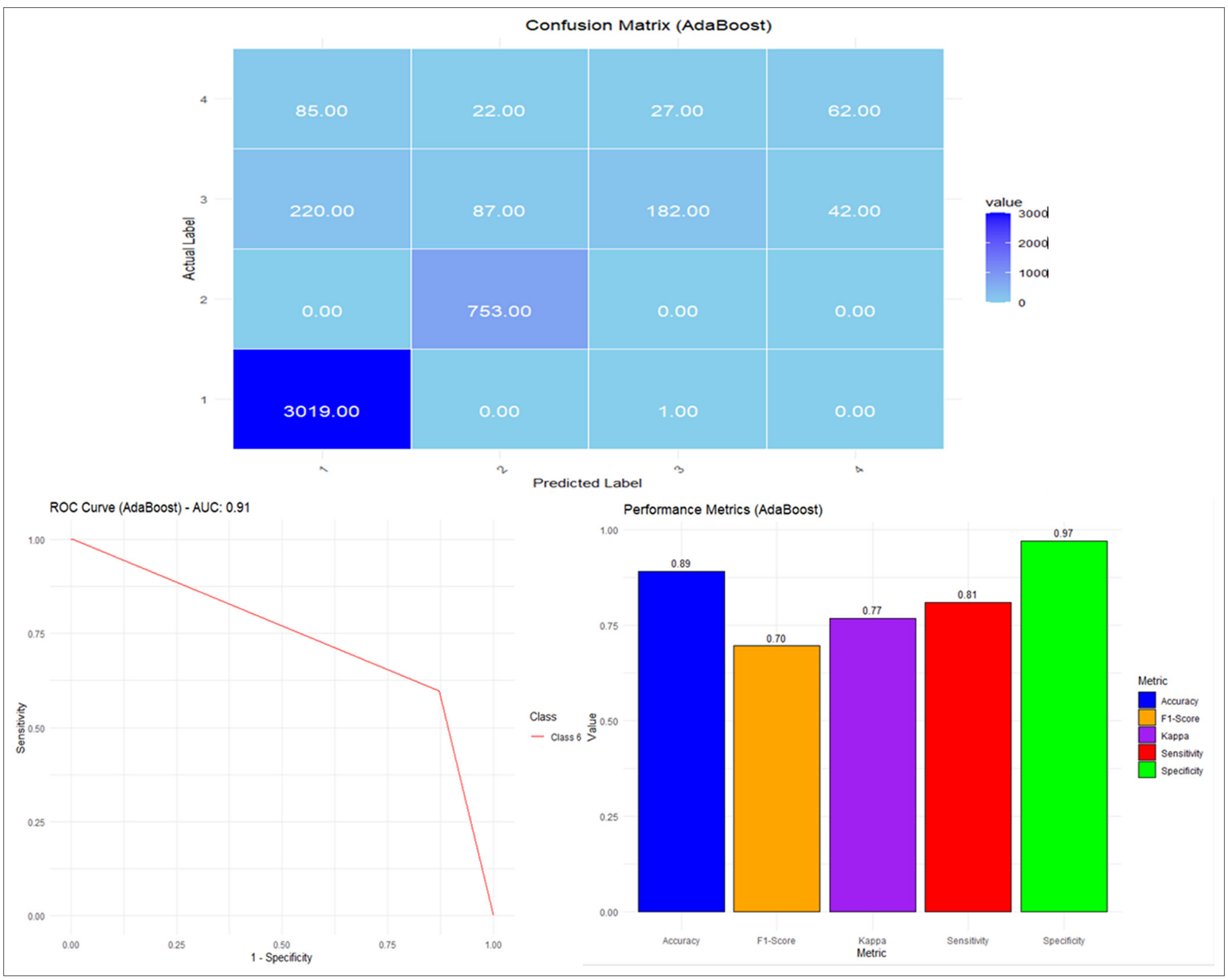

Figure 3: Graphs depicting different performance metrics for Adaboost algorithm in the classification of blood pressure levels. ROC: Receiver operating characteristic, AUC: Area under the curve.

DISCUSSION

The study assesses the predictive power of various advanced machine algorithms over the conventional multinomial logistic model in multiclassifying BP levels. In particular, the results of the Naïve Bayes, AdaBoost, FNN, and LSTM algorithms were compared with the benchmark multinomial logistic model using standard performance evaluation metrics [Table 1 and Figures 3-7].

Table 1 provides a comprehensive comparison of the performance metrics across the different ML models evaluated in this study. It includes the training and test accuracies, F1 scores, kappa statistics, sensitivities, specificities, and AUC scores, along with the confusion matrices for each model. From Table 1, all models achieved similar test accuracies, hovering around 89%. The FNN demonstrated the highest training accuracy at 93.33%,

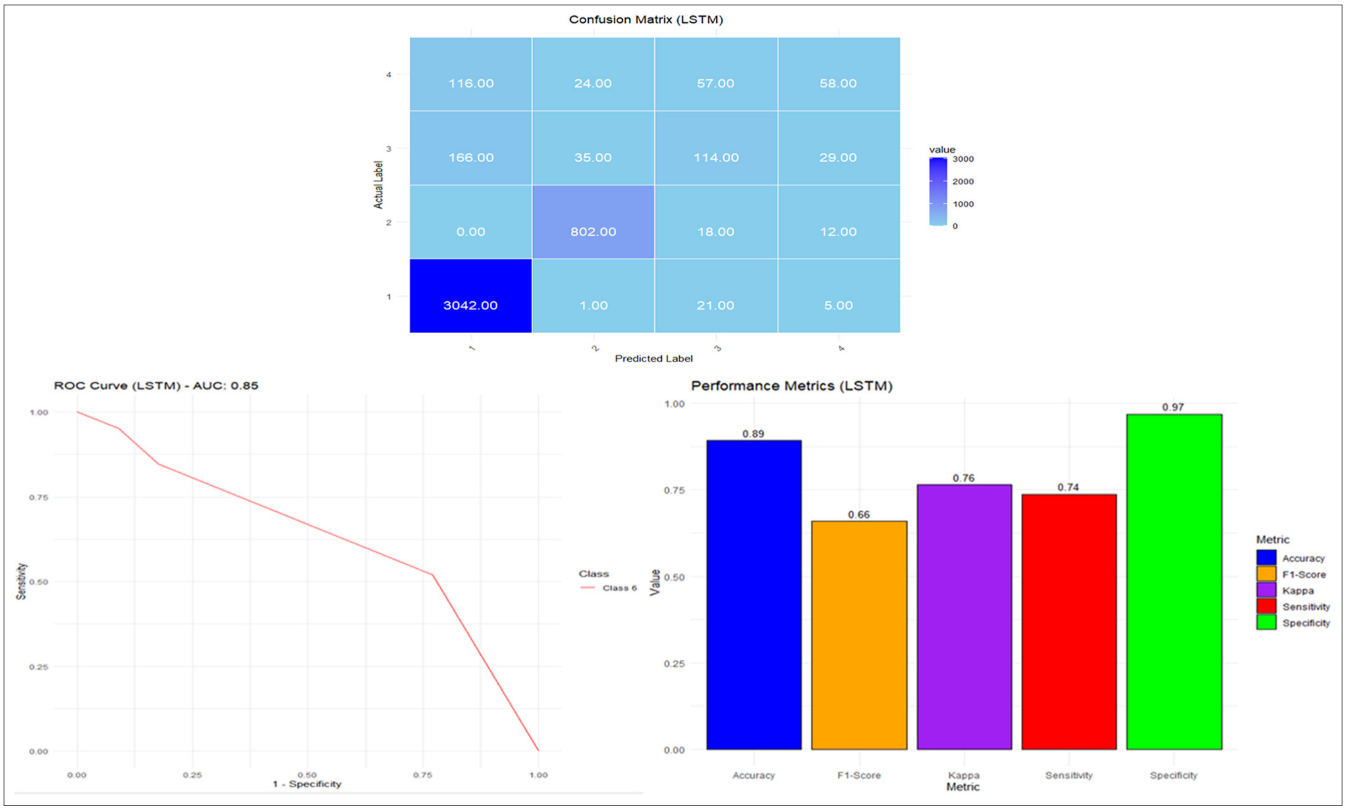

Figure 4: Graphs depicting different performance metrics for long short-term memory algorithm in the classification of blood pressure levels.

Table 1: Comparative performance metrics of various machine learning models.

Model	Training accuracy	Test accuracy	F1-score	Kappa	Sensitivity	Specificity	AUC score	Confusion matrix
AdaBoost	83.43	89.24	0.7	0.77	0.81	0.97	0.91	Class 1: 3019, Class 2: 753, Class 3 and 4: Misclassifications
LSTM	92.8	89.24	0.66	0.76	0.74	0.98	0.85	Class 1: 3042, Class 2: 802, Class 3 and 4: Slightly better than AdaBoost
FNN	93.3	89.47	0.65	0.77	0.72	0.98	0.84	Class 1: 3030, Class 2: 812, Class 3 and 4: Like AdaBoost and LSTM
Naïve Bayes	90	89.07	0.69	0.79	0.78	0.95	0.88	Class 1: 3003, Class 2: 850, Class 3: Better handling, Class 4: Some misclassifications
Logistic regression	83.21	89.07	0.68	0.77	0.79	0.97	0.89	Class 1: 2970, Class 2: 832, Class 3: 144, Class 4: 101

FNN: Feedforward neural network, LSTM: Long short-term memory, AUC: Area under the curve

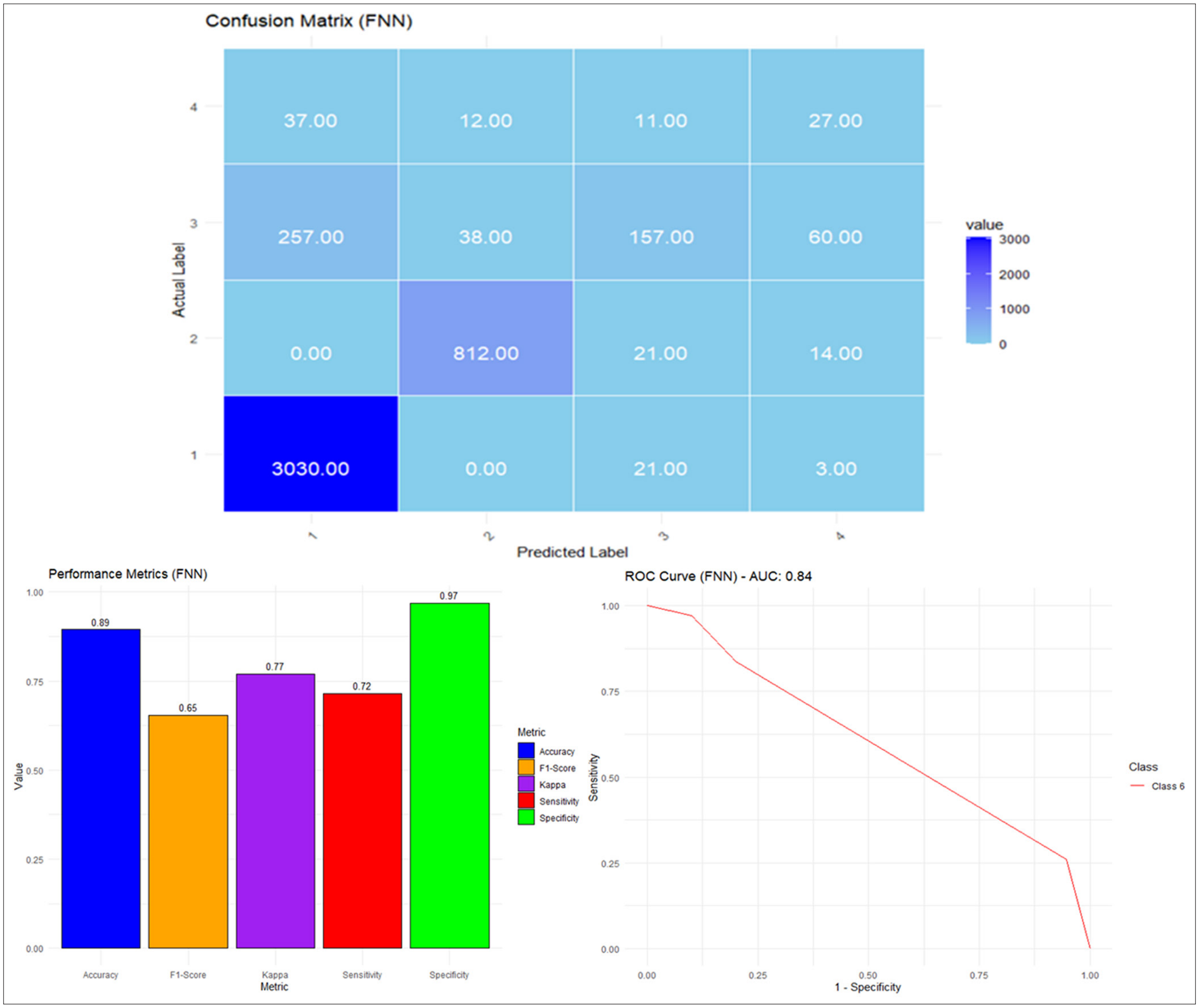

Figure 5: Graphs depicting different performance metrics for feedforward neural network algorithm in the classification of blood pressure levels.

indicating robust learning from the training data. However, its test accuracy (89.47%) suggests slight overfitting. In contrast, the LSTM model, when also exhibiting high training accuracy (92.8%), showed a balanced test accuracy (89.24%) and a relatively stable performance across different classes. Notably, the multiclass approach adopted from the American College of Cardiology and American Heart Association[9] has shown evidence of effectiveness.

AdaBoost, with a slightly lower training accuracy (83.43%), managed to perform equally well on the test set (89.24%), indicating its efficacy in handling class imbalance due to its boosting mechanism. The Naïve Bayes model, having the lowest test accuracy (89.07%), still showed respectable performance metrics, particularly in terms of specificity (0.97) and AUC score (0.88). The multinomial logistic model, used as a baseline, performed admirably with an accuracy of 89.07% and an AUC score of 0.89, demonstrating its capability as a reliable and interpretable model for this classification task in agreement with empirical literature.[7,9,12] The F1-Scores, which balance precision and recall, were notably higher for AdaBoost (0.70) and Naïve Bayes (0.69), compared to LSTM (0.66) and FNN (0.65), indicating better performance in managing the trade-off between false positives and false negatives. The kappa statistics, which measured the agreement between observed and predicted classifications, were also consistently high across models, with the highest being for Naïve Bayes (0.79).

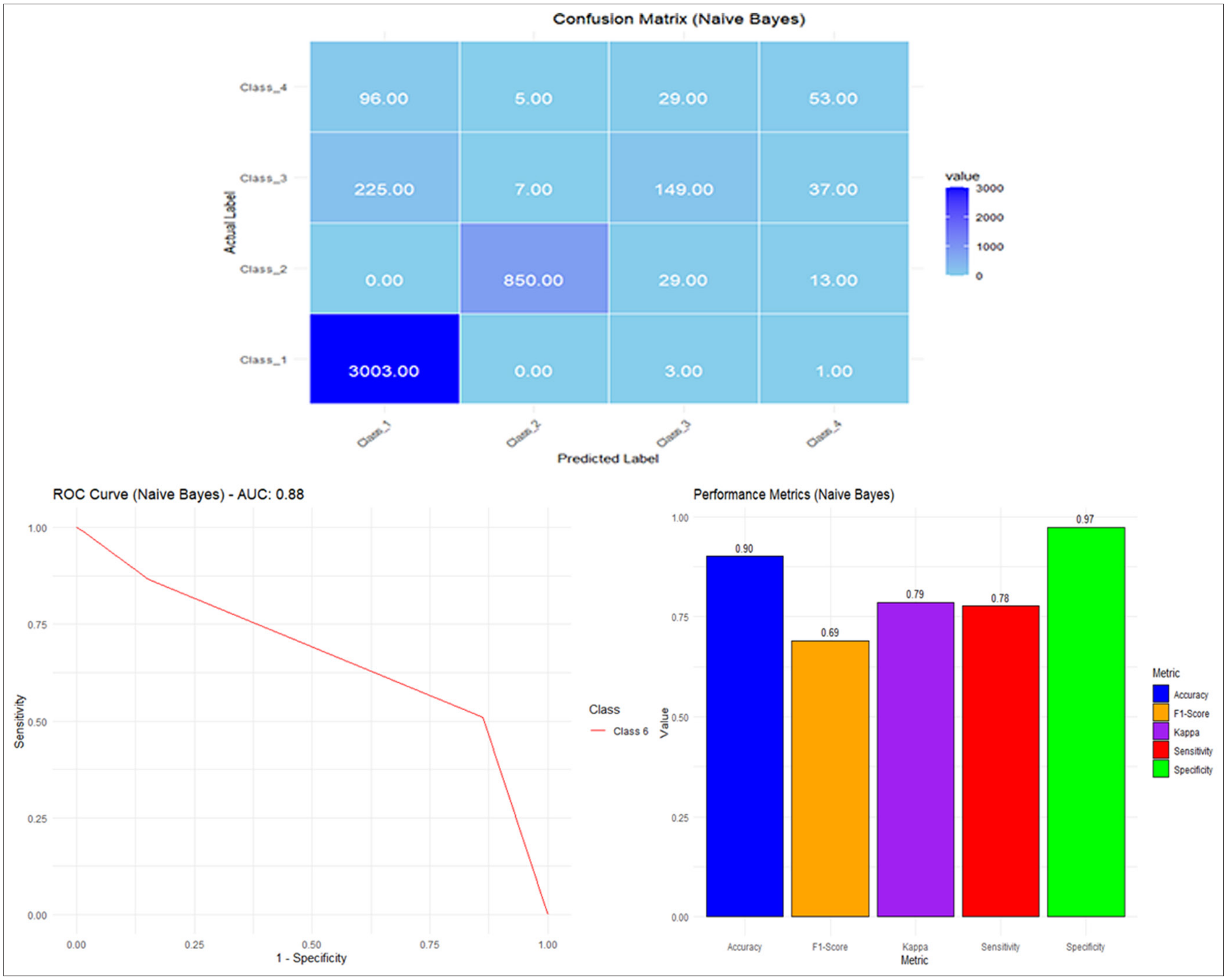

Figure 6: Graphs depicting different performance metrics for Naïve Bayes algorithm in the classification of blood pressure levels.

Overall, the table highlights that while FNN achieved the highest accuracy, the choice of model can depend on specific requirements such as the need for interpretability, computational efficiency, and the ability to handle imbalanced data. AdaBoost and Naïve Bayes provided balanced performance across metrics, whereas LSTM excelled in learning complex patterns, and logistic regression offered a solid baseline with consistent results. The LSTM model demonstrated superior generalization capabilities compared to the FNN model. Despite slight overfitting signs in the LSTM's training accuracy, it maintained strong performance across the test set, indicating robust generalization. Both LSTM and FNN models showed a consistent decrease in training loss. However, LSTM continued to improve steadily, whereas FNN displayed significant variability in its validation accuracy and loss, highlighting LSTM's superior learning efficiency. The LSTM model is better suited to capture temporal dependencies and complex patterns in the data, making it a more effective choice for datasets with such characteristics. The FNN, when showing high accuracy, struggled with the variability in the validation metrics, indicating potential overfitting.

Comparatively, the AdaBoost gives a high test accuracy and balanced performance with slightly lower sensitivity for rare classes. LSTM gives a high training accuracy with signs of slight overfitting and performs well across most classes. FNN gives a strong overall outperformance with the highest accuracy and balanced predictions. It is worth noting that the use of SMOTE uniquely enhances robustness and superiority in the output compared to previous studies.[8,11,13] Naïve Bayes is effective for common classes but less so for rare ones. The multinomial logistic produces a balanced performance, serving as a reliable baseline. Thus, the FNN emerges as the best model due to its high accuracy, balanced performance,

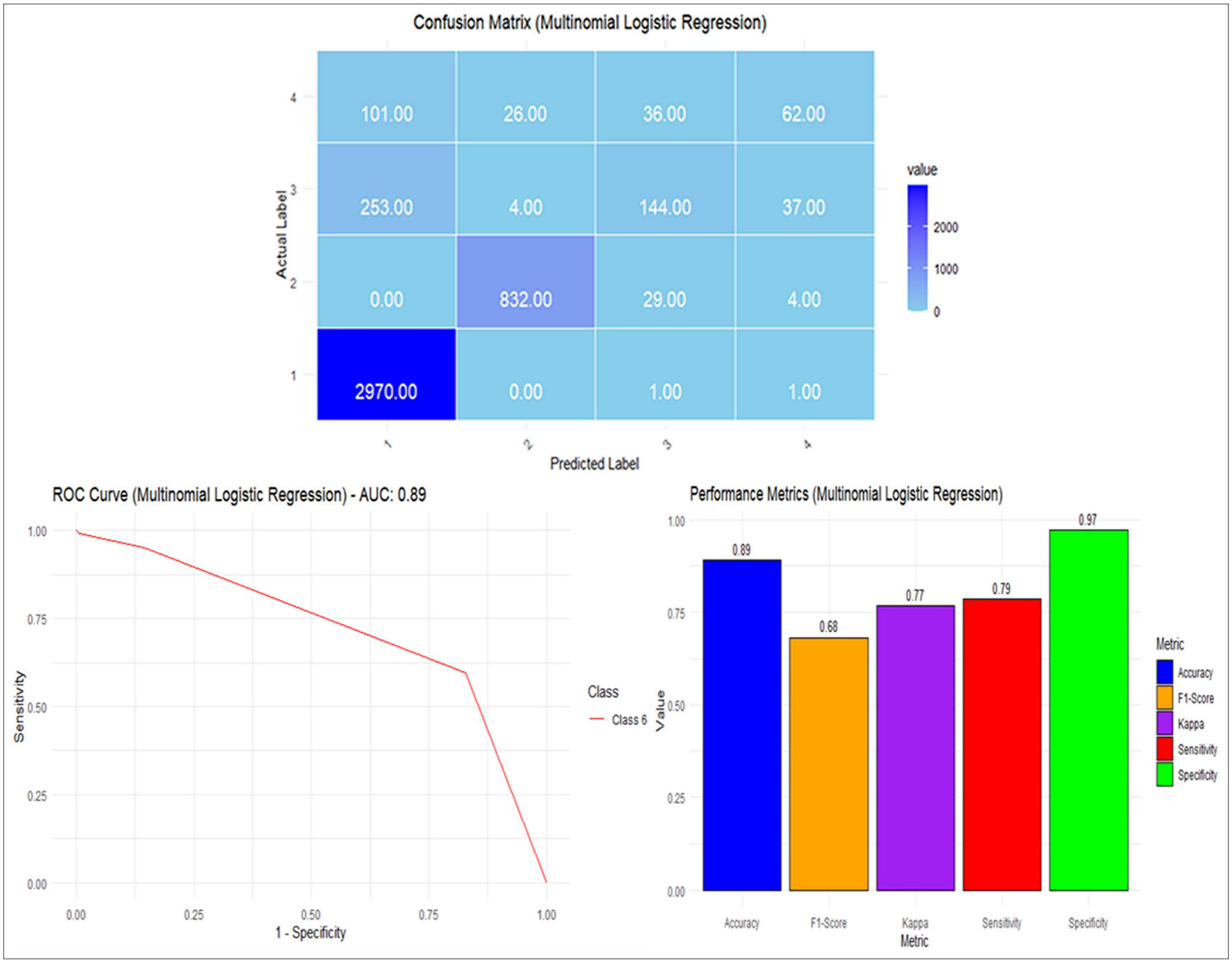

Figure 7: Graphs depicting different performance metrices for multinomial logistic algorithm in the classification of blood pressure levels.

and effective handling of class imbalance, making it suitable for predicting BP levels in this dataset. Despite its high performance, practitioners should consider the complexity and training stability of LSTM for datasets with similar characteristics.

CONCLUSION

Our study demonstrates that advanced ML techniques can significantly enhance the classification of BP levels, a crucial factor in preventing, detecting, and managing hypertension. Among the models tested, the FNN achieved the highest test accuracy (89.47%), making it highly promising for practical use in predicting BP levels, followed by LSTM, which gives 89.24%. AdaBoost is effective in handling class imbalance, and Naïve Bayes offers competitive performance despite its simplicity. These results represent a notable improvement over previous research, which typically reported median accuracy rates in the 80-85% range. The use of SMOTE to address class imbalance was instrumental in reaching these high and robust accuracy levels, highlighting the importance of balanced datasets in ML applications. The study underscores the importance of selecting appropriate ML models based on specific criteria such as accuracy, interpretability, and the ability to manage class distribution imbalances. The findings have significant implications for improving hypertension screening programs and developing predictive tools that facilitate early detection and prevention of hypertension. It reveals that knowing an individual's age, weight, height, gender, smoking habit, alcohol consumption, and fitness level are useful in predicting their BP levels. In this regard, the approach will enhance early detection or prevention with a view to saving lives. Further studies could focus on advanced clinical trials to validate these models in real-world settings, to ensure further applicability and effectiveness across different populations.

Acknowledgment: The authors acknowledge London Metropolitan University for internally funding this project via the Rescaling Fund 2023/2024 intervention.

Authors' contributions: JI and BNV: Conceptualized, designed, and introduced the study; IGO and IAA: Wrote the methods section; IOO and JI: Conducted the analysis and presented and discussed the results; SA: Recruited as a research assistant to assist in drafting the manuscript. All authors contributed to the full manuscript and unanimously agreed that the manuscript should be submitted for publication.

Ethical approval: Ethical approval is not required as the study uses secondary dataset that does not require human participants, no data subject, and no identifiable human characteristics.

Declaration of patient consent: Patient's consent not required as patient identity is not disclosed or compromised.

Financial support and sponsorship: London Metropolitan University internally funded this project via the Rescaling Fund 2023/2024 intervention.

Conflicts of interest: There are no conflicts of interest.

Availability of data and material: The cleaned dataset is available on request from the corresponding author.

REFERENCES

1. NHS blood pressure test: NHS.UK; 2024. Available from: https://www.nhs.uk/conditions/blood-pressure test/#:~:text=blood%20pressure%20is%20measured%20in,is%20pushed%20around%20your%20heart [Last accessed on 2025 Apr 21].
2. High blood pressure dangers: Hypertension's effects on your body. Mayo Clinic; 2023. Available from: https://www.mayoclinic.org/diseases-conditions/high-blood-pressure/in-depth/high-blood-pressure/art-20045868#:~:text=high%20blood%20pressure%20forces%20the,heart%20attack%20and%20heart%20failure [Last accessed on 2025 Apr 21].
3. Campbell E, Macey E, Finning K, Shine C. Risk factors for undiagnosed high blood pressure in England. Office for National Statistics; 2023. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandwellbeing/articles/riskfactorsforundiagnosedhighbloodpressureinengland/2015to2019#:~:text=1.-,main%20points,million%20adults%20with%20undiagnosed%20hypertension [Last accessed on 2025 Apr 21].
4. Jessica G. How much does the general public know about their blood pressure? - Health innovation east. Health Innovation East; 2023. Available from: https://healthinnovationeast.co.uk/blood-pressure-awareness/#:~:text=around%20one%20in%20three%20adults,and%20don't%20know%20it [Last accessed on 2025 Apr 21].
5. Katella K. 17 Things to know about high blood pressure. Yale Medicine; 2023. Available from: https://www.yalemedicine.org/news/high-blood-pressure-hypertension#:~:text=It's%20estimated%20that%20almost%20half,at%20risk%20for%20the%20condition [Last accessed on 2025 Apr 21].
6. Falkner B, Gidding SS, Portman R, Rosner B. Blood pressure variability and classification of prehypertension and hypertension in adolescence. Pediatrics 2008;122:238-42.
7. Saloni R, Sharma K, Gupta AK. Classification of high blood pressure persons vs normal blood pressure persons using voice analysis. IJ Image Graph Signal Process 2014;1:47-52.
8. Dong Y, Song Y, Zou Z, Ma J, Dong B, Prochaska JJ. Updates to pediatric hypertension guidelines: Influence on classification of high blood pressure in children and adolescents. J Hypertens 2019;37:297-306.
9. Wu S, Song Y, Chen S, Zheng M, Ma Y, Cui L, *et al*. Blood pressure classification of 2017 associated with cardiovascular disease and mortality in young Chinese adults. Hypertension 2020;76:251-8.
10. Shrivastava A, Chakkaravarthy M, Shah MA. A new machine learning method for predicting systolic and diastolic blood pressure using clinical characteristics. Healthc Anal 2023;4:100219.
11. Whelton PK, Carey RM, Aronow WS, Casey DE Jr., Collins KJ, Himmelfarb CD, *et al*. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APHA/ASH/ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: A report of the American college of cardiology/American heart association task force on clinical practice guidelines. Hypertension 2018;71:e1269-324.
12. Jeffrey S. New ACC/AHA hypertension guidelines make 130 the new 140. Medscape Medical News; 2017. Available from: https://www.medscape.com/viewarticle/888560?form=fpf, [Last accessed on 2025 Apr 21].
13. Irmansyah M, Madona E, Nasution A, Putra R. Low cost heart rate portable device for risk patients with IOT and warning system. In: 2018 International conference on applied information technology and innovation (ICAITI). United States: IEEE; 2018. p. 46-9.
14. Cherrier S, Deshpande V. From BPM to IOT. In: Business process management workshops: BPM 2017 international workshops, Barcelona, Spain, September 10-11, 2017, Revised Papers. Cham: Springer International Publishing; 2018. p. 310-8.
15. Abba S, Garba AM. An IOT-based smart framework for a human heartbeat rate monitoring and control system. In: Proceedings. Vol. 42. Switzerland: MDPI; 2019. doi: https://doi.org/10.3390/ecsa-6-06543.
16. Khan MN, Noor DF, Pantho MJ, Abtahi TS, Parveen F, Bhuiyan MI. A low cost optical sensor based heart rate monitoring system. In: 2013 International conference on informatics, electronics and vision (ICIEV). United States: IEEE; 2013. p. 1-5.
17. Sree KN, Rao GM. IOT based low cost and ageing healthcare monitoring system. Int J Eng Adv Technol 2020;9:6-14.
18. Murali M, Bhargava M, Snehaa G, Anand A, Haque A, Sarobin VR. Data analytics on IOT-based health monitoring system. Int J Recent Technol Eng 2019;8:220-3.
19. Al-Sheikh MA, Ameen IA. Design of mobile healthcare monitoring system using IOT technology and cloud computing. In: IOP conference series: Materials science and engineering. Vol. 881. United Kingdom: IOP Publishing; 2020. p. 012113.
20. Banerjee PS, Karmakar A, Dhara M, Ganguly K, Sarkar S. A novel method for predicting bradycardia and atrial fibrillation using fuzzy logic and arduino supported IOT sensors. Med Novel Technol Devices 2021;10:100058.
21. Islam M, Rahaman A, Islam R. Development of smart

healthcare monitoring system in IOT environment. SN Comput Sci 2020;1:185.
22. Tamilselvi V, Sribalaji S, Vigneshwaran P, Vinu P, GeethaRamani J. IOT based health monitoring system. In: 2020 6th International conference on advanced computing and communication systems (ICACCS). United States: IEEE; 2020. p. 386-9.
23. Acharya AD, Patil SN. IOT based health care monitoring kit. In: 2020 Fourth international conference on computing methodologies and communication (ICCMC). United States: IEEE; 2020. p. 363-8.
24. Rish I. An empirical study of the naïve Bayes classifier. In: Proceedings of the IJCAI 2001 workshop on empirical methods in artificial intelligence. Vol. 3. 2001. p. 41-6.
25. Arif MH, Li J, Iqbal M, Liu K. Sentiment analysis and spam detection in short informal text using learning classifier systems. Soft Comput 2018;22:7281-91.
26. Harzevili NS, Alizadeh SH. Mixture of latent multinomial naive Bayes classifier. Appl Soft Comput 2018;69:516-27.
27. McCallum A, Nigam K. A comparison of event models for naïve Bayes text classification. In: Proceedings in AAAI-98 workshop on learning for text categorization. Vol. 752. 1998. p. 41-8.
28. Hosmer DW Jr., Lemeshow S, Sturdivant RX. Applied logistic regression. United States: John Wiley and Sons; 2013. Available from: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118548387.fmatter?msockid=267a345f05c26fd62fe7207c04bb6e41. [Last accessed on 2025 Apr 21].
29. Lenz ST, Agresti A. Categorical data analysis. In: Statistical Papers. Vol. 57. Springer: London; 2013. p. 849-50.
30. Jayaprakash D, Kanimozhiselvi CS. Multinomial logistic regression method for early detection of autism spectrum disorders. Meas Sens 2024;33:101125.
31. Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. J Comput Syst Sci 1997;55:119-39.
32. Ho JC, Sotoodeh M, Zhang W, Simpson RL, Hertzberg VS. An AdaBoost-based algorithm to detect Hospital-acquired pressure injury in the presence of conflicting annotations. Comput Biol Med 2024;168:107754.
33. Schapire RE, Freund Y. Boosting: Foundations and algorithms. Kybernetes 2013;42:164-6.
34. Shashidhar R, Patilkulkarni S, Puneeth SB. Audio visual speech recognition using feed forward neural network architecture. In: 2020 IEEE international conference for innovation in technology (INOCON). United States: IEEE; 2020. p. 1-5.
35. Nielsen MA. Neural networks and deep learning. San Francisco, CA, USA: Determination Press; 2015. Available from: https://books.google.co.in/books?id=STDBswEACAAJ. [Last accessed on 2025 Apr 21].
36. Fine TL. Feedforward neural network methodology. Germany: Springer Science and Business Media; 2006. Available from: https://books.google.co.in/books?id=s-PlBwAAQBAJ. [Last accessed on 2025 Apr 21].
37. Razavi S, Tolson BA. A new formulation for feedforward neural networks. IEEE Trans Neural Netw 2011;22:1588-98.
38. Sharma S, Sharma S, Athaiya A. Activation functions in neural networks. Towards Data Sci 2017;6:310-6.
39. Ben Braiek H, Khomh F. Testing feedforward neural networks training programs. ACM Trans Soft Eng Methodol 2023;32:1-61.
40. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997;9:1735-80.
41. Van Houdt G, Mosquera C, Nápoles G. A review on the long short-term memory model. Artif Intell Rev 2020;53:5929-55.
42. Goodfellow I, Bengio Y, Courville A, Bengio Y. Deep learning. Cambridge: MIT press; 2016. Available from: https://mitpress.mit.edu/9780262035613/deep-learning/. [Last accessed on 2025 Apr 21].
43. Jiang C, Chen Y, Chen S, Bo Y, Li W, Tian W, *et al*. A mixed deep recurrent neural network for MEMS gyroscope noise suppressing. Electronics 2019;8:181.