International Journal on Science and Technology (IJSAT)

E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

Predicting Hypertension Through Lifestyle Analytics using Machine learning

Palak Shahr[1], Bhargvi Patel[2]

[1]Research Scholar, [2]Assistant Professor
[1,2]Department of Computer Science & Engineering, KIRC (affiliated by GTU), Gujarat, India.

Abstract:
Hypertension is a major risk factor for cardiovascular diseases and remains a global health concern due to its high prevalence and asymptomatic nature. Early detection is crucial for timely intervention. This study applies machine learning techniques to predict hypertension using a large dataset of 174,982 records with demographic, clinical, and lifestyle features such as age, BMI, blood pressure, cholesterol, stress, physical activity, and dietary habits.
Data preprocessing included encoding categorical variables, normalization, and handling class imbalance. Models such as Random Forest, XGBoost, and Logistic Regression were trained and evaluated using accuracy, precision, recall, and F1-score. Feature importance analysis was conducted to identify key predictors.
Results show that age, BMI, stress, physical activity, and diet are strongly associated with hypertension. The models achieved high predictive performance, demonstrating their potential for large-scale risk assessment. This study highlights the role of machine learning in early detection and prevention, supporting data-driven healthcare strategies to reduce cardiovascular disease burden.

Keywords: Hypertension Prediction, Machine Learning, Cardiovascular Risk, Feature Importance, Early Detection.

I. INTRODUCTION

Hypertension, commonly known as high blood pressure, is one of the leading causes of cardiovascular diseases worldwide [1]. It is strongly associated with heart disease, stroke, and kidney failure, making it a major contributor to global morbidity and mortality [1]. A major concern is its asymptomatic nature, as many individuals remain unaware of the condition until serious complications arise, leading to delayed diagnosis and treatment [2].

Despite advancements in healthcare and diagnostic tools, a significant number of cases remain undiagnosed or untreated [3]. Traditional detection methods rely mainly on clinical blood pressure measurements, which may not fully reflect an individual's overall risk profile [4]. Important factors such as lifestyle, behavior, and genetic predisposition are often overlooked, highlighting the need for more comprehensive approaches [5].

Recent advancements in data availability and computational techniques have enabled the use of machine learning for disease prediction [6]. Machine learning models can analyze large datasets and identify complex patterns among demographic, clinical, and lifestyle factors, providing deeper insights into hypertension risk [7]. These models support both individual risk assessment and large-scale public health strategies [7].
Furthermore, machine learning helps in understanding the relationship between modifiable factors such as physical activity, diet, stress, and sleep, and non-modifiable factors like age and family history [8]. This enables early detection, personalized healthcare recommendations, and improved preventive

strategies. Additionally, such approaches can support policy-making and targeted screening programs [9].

Overall, integrating machine learning into hypertension research represents a shift toward proactive and preventive healthcare [10]. By identifying at-risk individuals early, it can help reduce the global burden of hypertension and associated diseases.

A. Research Problem

Hypertension is a serious yet often asymptomatic condition that significantly contributes to cardiovascular diseases worldwide [11]. Due to the absence of visible symptoms, many individuals remain undiagnosed until severe complications occur [12]. Although various factors such as age, BMI, cholesterol, smoking, alcohol consumption, diet, stress, and physical activity influence its development, current risk assessment methods remain limited [13]. Traditional approaches primarily rely on isolated clinical measurements like blood pressure and cholesterol, failing to capture the complex interactions between lifestyle and behavioral factors [14].

This limitation creates a gap in early identification of at-risk individuals, reducing the effectiveness of preventive healthcare strategies [14]. Existing studies often analyze risk factors independently, making it difficult to understand their combined impact on hypertension [15]. Additionally, many predictive tools lack scalability and interpretability, limiting their practical use in clinical and public health settings [16]. Machine learning offers a promising solution by enabling the analysis of large datasets and uncovering hidden patterns beyond traditional statistical methods [17]. However, challenges remain regarding the accuracy, scalability, and interpretability of these models. Without interpretability, such models may be perceived as "black boxes," restricting their adoption in real-world healthcare applications [18].

Therefore, the core research problem is the need to develop accurate, scalable, and interpretable machine learning models that integrate demographic, clinical, and lifestyle factors to predict hypertension risk. Addressing this gap can support early detection, personalized prevention strategies, and large-scale screening, ultimately reducing the global burden of hypertension and its associated complications.

B. objective

The main objective of this study is to develop and evaluate machine learning models for the early prediction of hypertension using a comprehensive dataset that includes demographic, clinical, and lifestyle factors. The study begins with data preprocessing and analysis, which involves cleaning the dataset, normalizing continuous variables, encoding categorical features, and handling class imbalance to ensure accurate and reliable predictions.

Another important objective is to identify the key factors influencing hypertension risk by analyzing variables such as body mass index, stress levels, sleep duration, cholesterol, and physical activity. The study also aims to develop and compare multiple machine learning models, including Logistic Regression, Random Forest, and XGBoost, and evaluate their performance using metrics such as accuracy, precision, recall, F1-score, and AUC.

Furthermore, the research seeks to examine the trade-off between model accuracy and interpretability to ensure that the results are both reliable and understandable for healthcare professionals. It also aims to translate predictive outcomes into practical applications by providing recommendations for lifestyle modifications and preventive strategies for high-risk individuals. Finally, the study intends to demonstrate that machine learning can serve as a scalable and cost-effective approach for hypertension risk prediction across large populations, supporting both individual healthcare and broader public health initiatives.

C. Research Gaps

Despite extensive research on hypertension, several gaps remain in the use of predictive modeling for early detection and prevention. First, many studies analyze risk factors such as age, BMI, and smoking individually, with limited efforts to integrate demographic, clinical, lifestyle, and socioeconomic variables into a unified predictive framework [19]. This restricts the ability to capture the multifactorial nature of hypertension.

Second, a significant portion of existing research relies on traditional statistical methods like logistic regression, which have limitations in identifying complex and nonlinear relationships among risk factors [20]. Advanced machine learning techniques, which are better suited for handling large-scale and multidimensional data, are still underutilized in hypertension studies.

Third, even when machine learning is applied, most studies focus on a single algorithm without conducting systematic comparisons. This makes it difficult to determine the optimal balance between predictive accuracy and model interpretability, both of which are essential for real-world healthcare applications.

Finally, there is a gap between predictive model development and practical implementation. Many studies do not emphasize the interpretation of results or their application in personalized healthcare and community-level prevention strategies [21]. Bridging this gap is essential to ensure that predictive analytics can effectively contribute to healthcare decision-making and policy development.

II. LITERATURE REVIEW

A. F. Turki used machine learning models such as SVR, neural networks, and random forests to predict blood pressure using pulse transit time, achieving higher accuracy than traditional methods. However, small dataset size, physiological variability, and computational complexity limit real-world applicability [22].

Sandhiya et al. applied models like LightGBM, XGBoost, CatBoost, and Random Forest on clinical and lifestyle data, achieving up to 92% accuracy with LightGBM. However, issues like limited dataset diversity, overfitting, and lack of interpretability restrict clinical adoption [23].

Sadiku et al. developed classification models using demographic and clinical data to predict hypertension, achieving high predictive accuracy and identifying key factors such as age and BMI. However, limited dataset size and lack of external validation affect generalizability [24].

D. Shimbo et al. explored the role of AI in hypertension diagnosis and management, highlighting its potential in prediction, monitoring, and personalized treatment. However, lack of diverse datasets, validation issues, and system integration challenges hinder real-world implementation [25].

K. K. Reddy et al. proposed a deep learning model for anomaly detection using wearable device data, enabling continuous and non-invasive monitoring. However, data quality issues, sensor noise, and lack of large-scale validation limit reliability [26].
A. Mondal and K. Hazra used machine learning models such as Decision Tree, Random Forest, SVM, and Logistic Regression for hypertension prediction, achieving up to 94.77% accuracy. However, dataset limitations, noise, and computational challenges affect scalability and generalization [27].

T. Mroz et al. used XGBoost on large EHR datasets to predict hypertension control with an AUC of around 0.75. However, lack of external validation, missing lifestyle factors, and limited interpretability restrict practical use [28].

S. H. Hwang et al. developed machine learning models using large national datasets from South Korea and Japan, showing strong predictive performance and generalizability. However, limited regional scope and reliance on self-reported data affect broader applicability [29].

A. B. Bada et al. applied deep learning with transfer learning for hypertension prediction, achieving improved performance over traditional models. However, small dataset size, overfitting risks, and lack of interpretability limit clinical acceptance [30].

H. Islam et al. used explainable AI with SHAP values to detect blood pressure abnormalities, improving both prediction and interpretability. However, small dataset size, computational complexity, and lack of real-time validation remain challenges [31].

S. Abbas et al. combined predictive feature engineering with deep learning, achieving higher accuracy than traditional models. However, small dataset size, overfitting, and lack of explainability limit real-world use [32].

J. S. Cho and J.-H. Park reviewed AI applications in hypertension, showing improved prediction, diagnosis, and personalized treatment. However, challenges such as data quality, bias, lack of validation, and ethical concerns limit adoption [33].

A. Singh and N. Verma explored AI-based approaches including machine learning, deep learning, and wearable technologies for hypertension management, achieving high accuracy levels. However, issues like data inconsistency, model bias, and limited generalizability remain [34].

A. Shrivastava et al. proposed a machine learning model using clinical and lifestyle data, where Random Forest achieved the best performance. However, reliance on limited datasets and lack of real-time validation restrict practical implementation [35].

B. U. Stephen et al. reviewed machine learning integration in mHealth systems, highlighting improved prediction and diagnosis capabilities. However, lack of real-world implementation, data quality issues, and system fragmentation limit effectiveness [36].

S. Montagna et al. applied machine learning on large-scale screening data, where Random Forest showed the highest accuracy in hypertension detection. However, data inconsistency, lack of longitudinal data, and absence of external validation limit reliability [37].

H. F. Golino et al. used machine learning models such as Random Forest and SVM to predict high blood pressure, improving early detection. However, small dataset size, lack of longitudinal data, and class imbalance affect performance [38].
E. Bisong et al. applied machine learning models in low- and middle-income countries, demonstrating scalable and cost-effective hypertension prediction. However, limited data quality, lack of diversity, and interpretability challenges restrict adoption [39].

III. PROPOSED FLOW OF THE RESEARCH

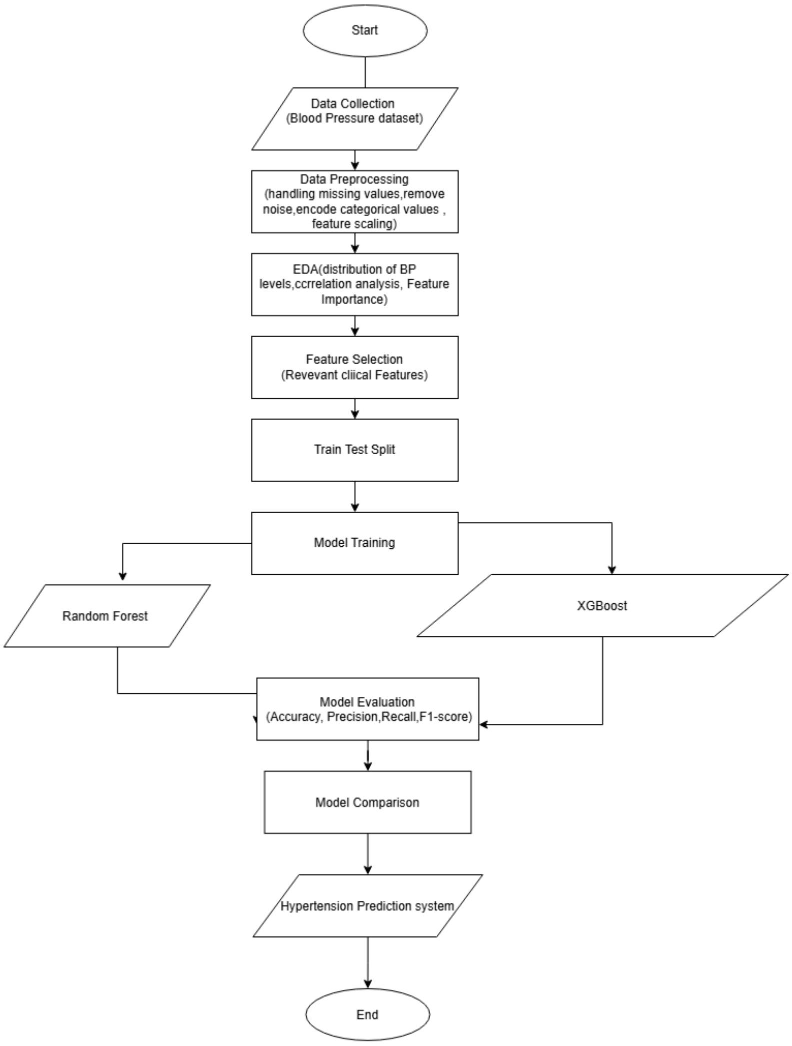

Fig 1: Proposed Flow of the Research

The proposed system begins with importing a comprehensive dataset that includes demographic, clinical, and lifestyle factors such as age, BMI, cholesterol levels, stress, sleep duration, smoking, and alcohol consumption. Data preprocessing is then performed to ensure quality and consistency by handling missing values, encoding categorical variables, and normalizing continuous features. This is followed by exploratory data analysis using visualization techniques to identify patterns, trends, and relationships among variables affecting hypertension risk.

The dataset is then divided into training and testing sets to evaluate model performance on unseen data. Machine learning models such as Random Forest, Gradient Boosting, and XGBoost are developed and trained due to their ability to handle complex and nonlinear relationships. These models are evaluated using metrics such as accuracy, precision, recall, F1-score, and confusion matrix to determine the most effective model for hypertension prediction.

To enhance interpretability, feature importance analysis is conducted to identify key contributing factors such as age, BMI, cholesterol, and lifestyle habits. The final results are interpreted to provide meaningful insights, which can be used for early detection, personalized healthcare recommendations, and development of preventive strategies to reduce hypertension risk.

The dataset used in this study includes a wide range of parameters categorized into demographic, clinical, and lifestyle factors. Demographic variables such as age, gender, country, education, and employment influence health awareness and access to healthcare. Clinical parameters include BMI, cholesterol, blood pressure, heart rate, glucose levels, and lipid profile, which directly impact cardiovascular health. Lifestyle factors such as smoking, alcohol consumption, physical activity, stress, diet, sleep duration, and family history further contribute to hypertension risk. The target variable classifies individuals based on hypertension status, enabling machine learning models to capture the multifactorial nature of the disease and improve prediction accuracy

IV. IMPLEMENTATION

A. Dataset Description

The dataset used in this study is stored in a CSV file named bloodpressure_updated.csv and was loaded into Python using the Pandas library. It was imported as a DataFrame using the read_csv() function, and initial inspection was performed using the head() function to verify its structure and contents. The dataset contains 22 variables, where each row represents an individual with a unique ID, including demographic, clinical, lifestyle, and hypertension-related information.

The dataset includes demographic variables such as age, gender, education, and employment status, which help explain variations in health patterns across populations. Clinical features such as BMI, heart rate, cholesterol levels, and blood pressure readings provide key indicators of cardiovascular health. Additionally, lifestyle factors like smoking, alcohol consumption, physical activity, diet, and sleep patterns are included, as they significantly influence hypertension risk.

The target variable represents the presence or absence of hypertension and is used for prediction in machine learning models. Overall, the dataset provides a comprehensive and multifactorial view of health, enabling accurate analysis and prediction of hypertension.

B. Data Pre-processing

The dataset bloodpressure_updated.csv was preprocessed to ensure it is clean, consistent, and suitable for training machine learning models. This process included handling missing values, encoding categorical variables, selecting relevant features, and applying feature scaling. These steps are essential to improve model accuracy and ensure reliable pattern learning from the data.

Missing values were first identified using appropriate functions and handled based on the type of data. Numerical features such as BMI were imputed using mean values to maintain data distribution, while categorical variables like gender were filled using the most frequent value (mode). This ensured completeness of the dataset and prevented errors during model training.

Categorical variables such as gender, education level, and employment status were converted into numerical form using Label Encoding, allowing machine learning algorithms to process them effectively. Feature selection was then performed using correlation analysis and domain knowledge to retain the most relevant variables such as age, BMI, cholesterol, blood pressure, and lifestyle factors like smoking, stress, and physical activity.
Finally, feature scaling was applied using standardization to ensure all variables contribute equally during model training. This transformation adjusts the data to have a mean of 0 and standard deviation of 1, improving model stability and performance by preventing features with larger values from dominating the learning process.

International Journal on Science and Technology (IJSAT)

E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

C. Data Visualization and Exploratory Analysis

Data visualization and exploratory analysis were performed to understand patterns, trends, and relationships within the dataset. Various graphical techniques such as box plots, heatmaps, scatter plots, histograms, and pie charts were used to examine how demographic, clinical, and lifestyle factors are associated with hypertension. These visualizations help in identifying key variables and understanding their influence on hypertension risk.

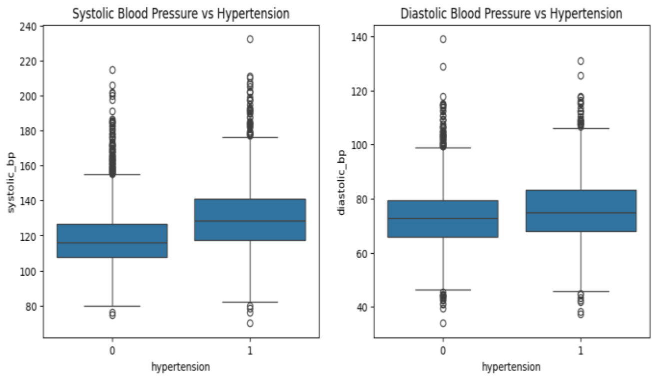

Fig 2: Systolic/Diastolic blood pressure vs hypertension

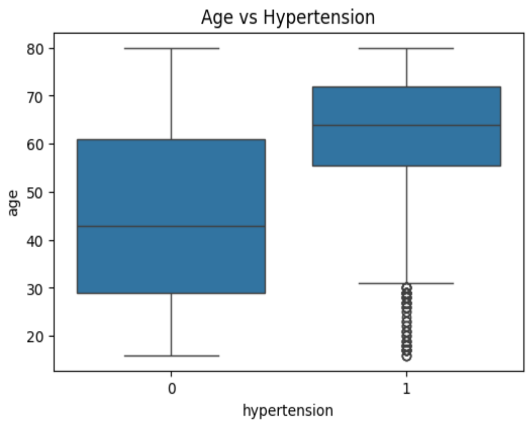

Fig 3: Age vs Hypertension

The analysis showed that both systolic and diastolic blood pressure values are significantly higher in individuals with hypertension compared to non-hypertensive individuals, confirming their importance as strong predictive features. Age was also found to have a strong relationship with hypertension, with older individuals showing a higher prevalence. Scatter plot analysis further indicated that systolic blood pressure tends to increase with age, reinforcing the role of aging as a major risk factor.

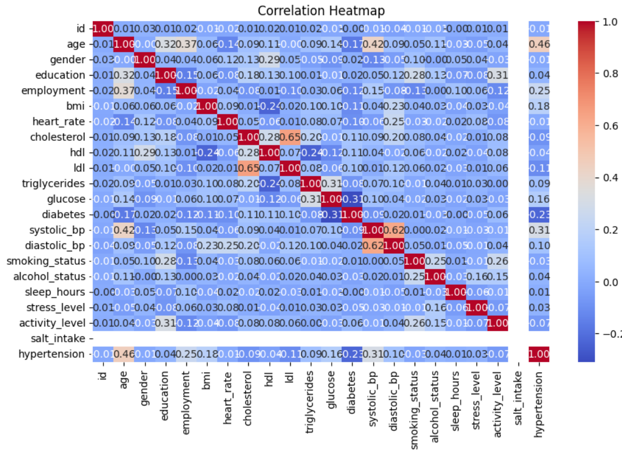

Fig 4: Correlation Heatmap

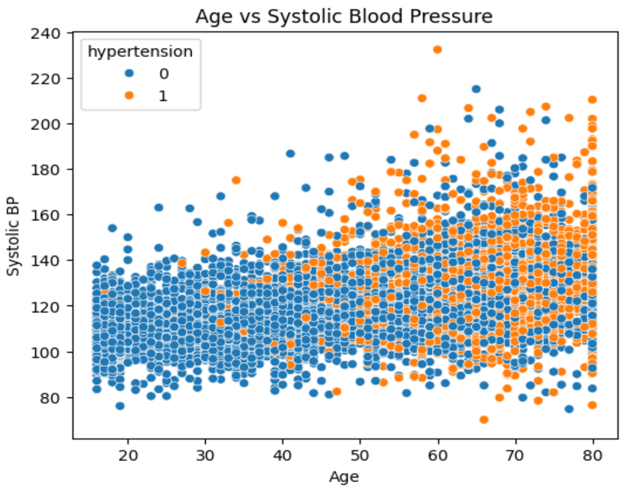

Fig 5: Pie chart illustration for the distribution of hypertension cases across different age groups

Correlation analysis revealed that age, systolic blood pressure, and diastolic blood pressure have strong positive relationships with hypertension, while variables such as cholesterol and lipid profiles show moderate associations. Lifestyle factors like physical activity, smoking, and stress also contribute to hypertension risk, although their correlations are relatively weaker compared to clinical factors.

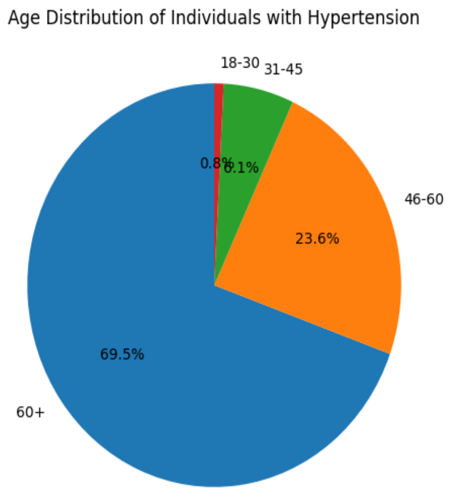

Fig 6: BMI distribution among hypertensive individuals

Additionally, the distribution analysis showed that most hypertensive individuals belong to older age groups, particularly above 60 years, and fall within higher BMI ranges, indicating overweight and obesity as significant risk factors. Overall, the exploratory analysis highlights that hypertension is influenced by a combination of demographic, clinical, and lifestyle variables, supporting the use of machine learning models for accurate prediction.

D. Model Development

After preprocessing and scaling, machine learning models were developed to predict hypertension based on demographic, clinical, and lifestyle features. Ensemble learning methods, specifically Random Forest and XGBoost, were selected due to their ability to handle complex and nonlinear relationships in data. The dataset was split into training (80%) and testing (20%) sets using the train-test split method to evaluate model performance on unseen data.

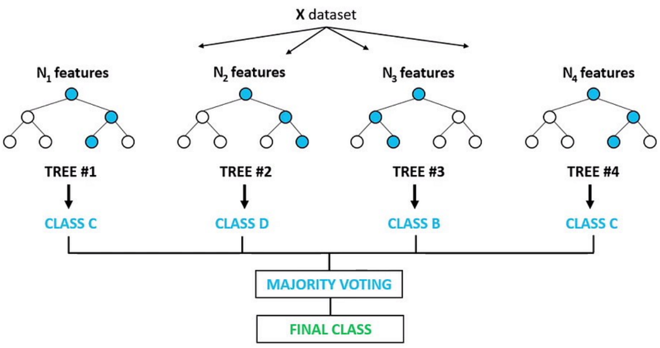

Fig 7: Random Forest Architecture

The Random Forest model was implemented as an ensemble of multiple decision trees, where each tree is trained on random subsets of data and features to reduce overfitting and improve stability. It was configured with 200 trees and a maximum depth of 10 to balance accuracy and complexity. This model is effective in healthcare prediction tasks as it can handle large datasets and provide insights into feature importance while maintaining high predictive performance.

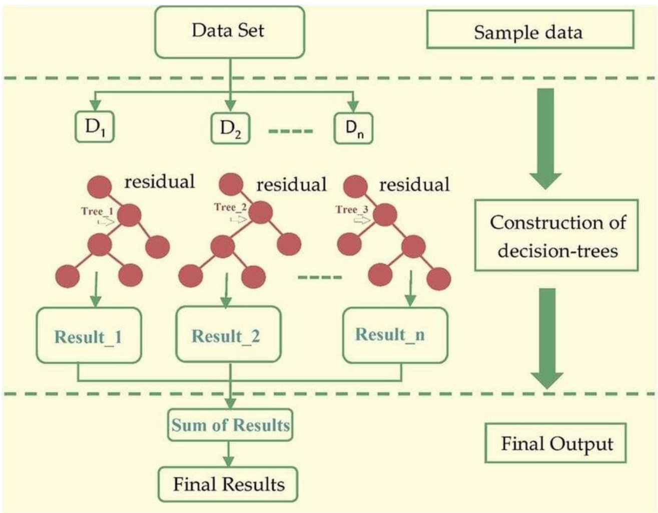

Fig 8: XGBoost Architecture

The XGBoost model was also implemented as a powerful boosting algorithm that builds trees sequentially, where each new tree corrects the errors of previous ones. It was configured with 200 estimators, a learning rate of 0.05, and a maximum depth of 6 to ensure controlled learning and prevent overfitting. XGBoost uses gradient boosting and regularization techniques, making it highly efficient and accurate for classification problems such as hypertension prediction.

V. RESULTS AND OBSERVATIONS

The performance of the proposed hypertension prediction system was evaluated using two ensemble machine learning models, Random Forest and XGBoost. Both models were trained and tested on the processed dataset using standard classification metrics such as accuracy, precision, recall, and F1-score. The results showed that both models performed similarly, with Random Forest achieving an accuracy of 79.67% and XGBoost achieving 79.35%, indicating comparable predictive capability .

A detailed class-wise analysis revealed that both models performed better in predicting non-hypertensive cases (Class 0) compared to hypertensive cases (Class 1). For Class 0, both models achieved high precision (0.83) and recall (around 0.89–0.90), indicating strong ability to correctly identify healthy individuals. However, for Class 1, the performance was lower, with Random Forest achieving precision of 0.69 and recall of 0.54, while XGBoost achieved slightly lower precision (0.68) but marginally higher recall (0.56). This indicates the presence of false negatives, where hypertensive individuals are misclassified as non-hypertensive, which is a critical concern in healthcare applications .

The comparison between the two models shows that Random Forest provides slightly better overall accuracy and stability due to its bagging approach, while XGBoost, based on boosting, demonstrates a slight advantage in detecting hypertensive cases. Despite parameter tuning, the performance difference between the two models remains minimal. Both models achieved similar F1-scores and balanced performance across evaluation metrics, confirming their effectiveness in hypertension prediction.

Table 1: Performance Comparison of Random Forest and XGBoost Models

Metric	**Random Forest**	**XGBoost**
Accuracy	**0.7967**	0.7936
Precision (Class 0)	0.83	0.83
Recall (Class 0)	**0.90**	0.89
F1-score (Class 0)	0.86	0.86
Precision (Class 1)	**0.69**	0.68
Recall (Class 1)	0.54	**0.56**
F1-score (Class 1)	0.61	0.61
Macro Avg F1-score	0.74	0.74
Weighted Avg F1-score	0.79	0.79

Overall, the results indicate that Random Forest is slightly more reliable for this study, while XGBoost offers competitive performance with better sensitivity to positive cases. However, both models require further improvement, particularly in reducing false negatives and improving recall for hypertensive cases. This highlights the need for future enhancements such as addressing class imbalance, incorporating additional features, and applying advanced modeling techniques to improve predictive accuracy and clinical applicability.

Overall, both models were trained on the dataset and evaluated using the testing set to assess their ability to generalize. Their implementation enables the system to capture complex patterns in health data and improve prediction accuracy, making them suitable for real-world healthcare applications

VI. CONCLUSIONS

Hypertension is a health determinant of great concern worldwide and a significant cause of cardiovascular diseases with millions of individuals being affected worldwide. People call it a silent killer as it has no early symptoms and is most commonly revealed only in the cases when the serious complications like heart disease, ischemia, or kidney failure have taken place. This highlights the importance of the early-detection and prevention.

This paper discusses a machine learning approach of identifying the risk of hypertension based on a combination of lifestyle, demographic, and clinical factors. This study is a multivariate one as opposed to the traditional diagnostic methods which make use of single measurements such as blood pressure. The variables it includes are age, gender, body mass index (BMI), cholesterol, heart rate, smoking status, alcohol intake, levels of stress, physical exercise, sleep patterns, and socioeconomic status.

Preprocessing methods such as treatment of missing values, categorical variables encoding, feature selection, and scaling were used to guarantee quality of the data. The results of the exploratory data analysis showed that hypertension is strongly correlated with such factors as age, obesity, and high cholesterol.

Two ensemble learning models, the Random Forest and the XGBoost, were tested and trained with an 80:20 train test split. The models had an accuracy of about 80-82. Random Forest gave consistent and understandable results, and XGBoost gave marginally better results with greater complexity.

The results indicate that there is the presence of multiple interacting factors that affect hypertension. Altogether, the research proves that machine learning could be successfully used to promote early diagnosis and risk assessment, which open the way to the better preventive care and the treatment plans according to the personal needs.

ACKNOWLEDGMENT

The heading of the Acknowledgment section and the References section must not be numbered.
Causal Productions wishes to acknowledge Michael Shell and other contributors for developing and maintaining the IEEE LaTeX style files which have been used in the preparation of this template. To see the list of contributors, please refer to the top of file IEEETran.cls in the IEEE LaTeX distribution.

REFERENCES:

[1] S. E. Kjeldsen, "Hypertension and cardiovascular risk: General aspects," *Pharmacological Research*, vol. 129, pp. 95-99, 2018.

[2] Â. Leite, "Chronic Illnesses: Varied Health Patterns and Mental Health Challenges," *Healthcare*, vol. 13, no. 12, pp. 36-47, 2025.

[3] G. Bilo, S. Brouwers, A. Kollias and R. J. McManus, "Bridging the gap between hypertension guidelines and real-life practice: strategies and technological tools for improving long-term population blood pressure control," *European Journal of Preventive Cardiology*, pp. 1-8, 2025.

[4] F.-Y. Chen, C.-W. Lee, Y.-J. Chen and Y.-H. Lin, "Pathophysiology and blood pressure measurements of hypertension in the elderly," *Journal of the Formosan Medical Association*, vol. 124, pp. 10-16, 2025.

[5] S. Kashum and G. Bahari, "Impact of Key Lifestyle Behaviors on Hypertension Control: Implications for Optimizing Patient Management," *Healthcare*, vol. 14, no. 1, pp. 1-8, 2026.

[6] M. Khalifa and M. Albadawy, "Artificial Intelligence for Clinical Prediction: Exploring Key

Domains and Essential Functions," *Computer Methods and Programs in Biomedicine Update*, vol. 5, pp. 100-148, 2024.

[7] L. Oneto, N. Navarin, A. Micheli and L. Pasa, "Informed machine learning for complex data," *Neurocomputing*, vol. 669, pp. 132-135, 2026.

[8] B. Chhetri, L. M. Goyal and M. Mittal, "How machine learning is used to study addiction in digital healthcare: A systematic review," *International Journal of Information Management Data Insights*, vol. 3, no. 2, pp. 100-107, 2023.

[9] C. Chakraborty, M. Bhattacharya, S. Pal and S.-S. Lee, "From machine learning to deep learning: Advances of the recent data-driven paradigm shift in medicine and healthcare," *Current Research in Biotechnology*, vol. 7, pp. 100-116, 2024.

[10] I. Naskinova, M. Kolev, D. Karova and M. Milev, "Machine Learning-Based Blood Pressure Prediction Using Cardiovascular Disease Data: A Comprehensive Comparative Study," *Electronics*, vol. 15, no. 2, pp. 310-312, 2026.

[11] S. Goorani, S. Zangene and J. D. Imig, "Hypertension: A Continuing Public Healthcare Issue," *Molecular Science*, vol. 26, no. 1, pp. 12-13, 2025.

[12] A. A. Karatza, S. Fouzas, D. Gkentzi and E. Kostopoulou, "Missed or Delayed Diagnosis of Heart Disease by the General Pediatrician," *Children*, vol. 12, no. 3, pp. 120-130, 2025.

[13] S. M. Mohamed, M. A. Shalaby, R. A. El-Shiekh and H. A. El-Banna, "Metabolic syndrome: risk factors, diagnosis, pathogenesis, and management with natural approaches," *Food Chemistry Advances*, vol. 3, pp. 100-102, 2023.

[14] M. Liu, T. Zhang, W. Guo and Y. Chang, "The association between blood pressure control and multiple self-management behaviors in elderly patients with hypertension: a study based on latent class analysis," *PubMed Central*, pp. 1-8, 2026.

[15] D. Osamika, B. S. Adelusi, M. C. Kelvin-Agwu and A. Y. Mustapha, "A Predictive Analytics Framework for Early Detection and Management of Cancer Using Multi-Source Health Data," *International Journal of Management and Organizational Research*, vol. 3, no. 1, pp. 223-232, 2024.

[16] E. Martinez-Ríos, L. Montesinos, M. Alfaro-Ponce and L. Pecchia, "A review of machine learning in hypertension detection and blood pressure estimation based on clinical and physiological data," *Biomedical Signal Processing and Control*, vol. 68, pp. 102-108, 2021.

[17] N. Tamascelli, A. Campari and T. Parhizkar, "Artificial Intelligence for safety and reliability: A descriptive, bibliometric and interpretative review on machine learning," *Journal of Loss Prevention in the Process Industries*, vol. 90, pp. 105-108, 2024.

[18] Z. Sadeghi, R. Alizadehsani, M. Akif and S. Kausar, "A review of Explainable Artificial Intelligence in healthcare," *Computers and Electrical Engineering*, vol. 118, pp. 109-113, 2024.

[19] I. D. Mienye, G. Obaido and N. Jere, "A survey of explainable artificial intelligence in healthcare: Concepts, applications, and challenges," *Informatics in Medicine Unlocked*, vol. 51, pp. 101-105, 2024.

[20] P. R. D. Riley, J. Ensor and K. I. E. Snell, "Importance of sample size on the quality and utility of AI-based prediction models for healthcare," *The Lancet Digital Health*, vol. 7, no. 6, pp. 100-108, 2025.

[21] N. Spezia, S. D. Rosis and S. Nuti, "Sense of Community in the context of disease prevention and health promotion: A scoping review of the literature," *BMC Public Health*, vol. 24, pp. 30-32, 2024.

[22] A. F. Turki, "Assessing the Efficacy of Various Machine Learning Algorithms in Predicting

Blood Pressure Using Pulse Transit Time," *Diagnostics- MDPI,* pp. 1-24, 2025.

[23] Sandhiya, T. D, S. S, S. K. K and S. K, "Enhanced Hypertension Disease Prediction using Machine Learning," Kongu Engineering College,, Perundurai, Erode, Tamil nadu, India, 2025.

[24] Sadiku, Ishola, Oladele, Adewuyi, Omotoso and Olorunmola, "Classification Model for Prediction of Hypertension," *Sule Lamido University Journal of Science & Technology,* vol. 11, no. 2, pp. 339-362, 2025.

[25] D. Shimbo, R. U. Shah, M. Abdalla, R. Agarwal and F. S. Ahmad, "Transforming Hypertension Diagnosis and Management in The Era of Artificial Intelligence: A 2023 National Heart, Lung, and Blood Institute (NHLBI) Workshop Report," *Hypertension,* vol. 42, pp. 36-46, 2025.

[26] K. K. Reddy, V. S. Kaza, R. M. Mohana, M. Alhameed and F. Jeribi, "Detecting anomalies in smart wearables for hypertension: a deep learning mechanism," *Frontiers in public health,* pp. 1-20, 2025.

[27] A. Mondal and K. Hazra, "Predictive Modeling for Hypertension Detection: A Machine Learning Approach," *International Journal of Novel Research And Development* , vol. 9, no. 5, pp. 1-9, 2024.

[28] T. Mroz, M. Griffin, R. Cartabuke, L. Laffin, G. Russo-Alvarez and G. Thomas, "Predicting hypertension control using machine learning," *PLOS ONE,* pp. 1-14, 2024.

[29] S. H. Hwang, H. Lee, J. Hyuk, Myeongcheol, A. Koyanagi and L. Smith, "Machine Learning–Based Prediction for Incident Hypertension Based on Regular Health Checkup Data: Derivation and Validation in 2 Independent Nationwide Cohorts in South Korea and Japan," *JOURNAL OF MEDICAL INTERNET RESEARCH,* vol. 26, pp. 1-14, 2024.

[30] A. B. Bada, Garko and D. Gabi, "HYPERTENSION PREDICTION USING DEEP LEARNING WITH TRANSFER LEARNING TECHNIQUES," *FUDMA Journal of Sciences,* vol. 8, no. 6, pp. 257-263, 2024.

[31] H. Islam, S. Iqbal and M. M. Hossain, "Blood pressure abnormality detection and interpretation utilizing explainable artificial intelligence," *Intelligent Medicine,* vol. 5, pp. 54-65, 2025.

[32] S. ABBAS, G. A. SAMPEDRO, M. KRICHEN and M. A. ALAMRO, "Effective Hypertension Detection Using Predictive Feature Engineering and Deep Learning," *IEEE,* pp. 55-69, 2024.

[33] J. S. Cho and J.-H. Park, "Application of artificial intelligence in hypertension," *Clinical Hypertension,* vol. 30, pp. 1-9, 2024.

[34] A. Singh and N. Verma, "Artificial Intelligence in Hypertension: A Futuristic Approach in Blood Pressure Management," *Hypertension Journal,* vol. 1, p. 10, 2023.

[35] A. Shrivastava, M. Chakkaravarthy and M. A. Shah, "A new machine learning method for predicting systolic and diastolic blood pressure using clinical characteristics," *Healthcare Analytics,* pp. 1-8, 2023.

[36] B. U. Stephen, B. C. Uzoewulu, P. M. Asuquo and S. Ozuomba, "Diabetes and hypertension MobileHealth systems: a review of general challenges and advancements," *Journal of Engineering and Applied Science,* vol. 70, pp. 1-40, 2023.

[37] S. Montagna, M. F. Pengo, S. Ferretti, C. Borghi, C. Ferri and G. Grassi, "Machine Learning in Hypertension Detection: A Study on World Hypertension Day Data," *Journal of Medical Systems* , pp. 1-10, 2023.

[38] H. F. Golino, L. S. d. B. Amaral, S. F. P. Duarte and CristianoMauro Assis Gomes, "Predicting Increased Blood Pressure Using Machine Learning," *Journal of Obesity,* pp. 1-12,

2022.

[39] E. Bisong, N. Jibril, P. Premnath, E. Buligwa and G. Oboh, "Predicting high blood pressure using machine learning models in lowand and middle-income countries," *BMC Medical Informatics and Decision Making*, vol. 24, pp. 1-22, 2024.