International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
ISSN:3048-7722
Volume 4, Issue 4, July-Aug-2026, PP: 01-5

An Intelligent Machine Learning Framework for Early Hypertension Risk Prediction Using Lifestyle and Clinical Health Indicators

M.Sai Manasa[1], L.srivani[2]
[1]Assistant Professor, Department of MCA, Megha Institute of Engineering and Technology for Women, Edulabad (Village), Ghatkesar (Mandal), Medchal District, Telangana –, India.
[2]MCA Student, Department of MCA, Megha Institute of Engineering and Technology for Women, Edulabad (Village), Ghatkesar (Mandal), Medchal District, Telangana –, India.

***Abstract* – Hypertension is one of the leading chronic diseases responsible for cardiovascular disorders and premature mortality across the world. Detecting individuals who are at risk during the early stages can significantly reduce severe health complications through timely medical intervention and lifestyle modification. This study presents a machine learning-based framework for predicting hypertension by analyzing demographic, clinical, and lifestyle-related attributes collected from a publicly available dataset. The proposed work investigates the performance of Decision Tree, Gradient Boosting, Extreme Gradient Boosting (XGBoost), and Random Forest classifiers. To improve predictive capability, are incorporated along with appropriate preprocessing and cross-validation strategies. The experimental findings indicate that feature optimization considerably enhances classification performance by reducing redundant information and improving model generalization. Among the evaluated algorithms, Gradient Boosting and XGBoost demonstrate superior prediction capability, while the hybrid learning approach further improves robustness and classification facilitating personalized hypertension management through data-driven decision-making.**

Keywords - **Hypertension Prediction, Machine Learning, Random Forest, Gradient Boosting, XGBoost, Decision Tree, Feature Extraction, PCA, LDA, Healthcare Analytics.**

I. INTRODUCTION

Since hypertension generally develops without noticeable symptoms during its early stages, many individuals remain undiagnosed until severe complications occur. Therefore, accurate and early identification of hypertension risk has become an important objective in modern healthcare systems.

have transformed the healthcare industry by enabling automated disease prediction using large-scale clinical datasets. Machine learning algorithms can recognize complex relationships among physiological measurements, demographic information, and lifestyle characteristics that may not be easily identified through conventional statistical approaches. Such predictive

Several health-related attributes including age, , family medical history, dietary habits, physical inactivity, smoking behavior, and glucose levels have been recognized as important indicators influencing hypertension development. Integrating these diverse variables within machine learning frameworks allows accurate prediction of disease risk while supporting personalized preventive healthcare strategies.

This research investigates multiple supervised machine learning algorithms for hypertension prediction using clinical and lifestyle data. Feature extraction techniques are employed to eliminate irrelevant information, improve computational efficiency, and enhance classification performance.methods for reliable hypertension prediction and supports their practical application in intelligent healthcare systems.

II. LITERATURE SURVEY

Machine learning has emerged as an effective technology for improving hypertension by analyzing large volumes of clinical and lifestyle-related data. Researchers have developed numerous predictive models that who are at high risk before severe cardiovascular complications develop. Compared with traditional statistical techniques, machine learning algorithms provide higher prediction accuracy, better scalability, and improved capability to capture complex relationships among multiple health parameters.

Several studies have explored the application of Support Vector Machine (SVM) for hypertension prediction because of its strong classification performance in high-dimensional medical datasets. SVM has demonstrated reliable results in distinguishing hypertensive and non-hypertensive patients, especially when combined with feature selection methods and balanced training data. Researchers have shown that integrating electronic health records with SVM-based models significantly enhances prediction accuracy while supporting clinical decision-making.

Ensemble learning minimize overfitting, and provide robust performance across diverse patient populations. Comparative studies consistently indicate that boosting algorithms outperform many conventional classifiers because of their ability to learn from previous prediction errors and optimize model performance iteratively.

Feature engineering has also become an important component of hypertension prediction systems. Dimensionality reduction techniques including Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are widely adopted to eliminate redundant attributes while preserving the most informative clinical features. These methods reduce computational complexity, improve model efficiency, and enhance prediction accuracy by removing irrelevant or noisy information from the dataset. Proper preprocessing, normalization, and feature optimization further contribute to developing more reliable predictive models.

combine the strengths of multiple algorithms to achieve better diagnostic performance. Hybrid models integrate classifiers variance, and increase robustness when handling heterogeneous healthcare data. Such integrated approaches have demonstrated excellent performance in identifying hypertension risk while maintaining high sensitivity and specificity across different patient groups.

Furthermore, advances in artificial intelligence, deep learning, and healthcare analytics continue to expand the scope of hypertension prediction. The integration of wearable sensor data, electronic medical records, physiological measurements, and lifestyle information enables the development of intelligent decision-support systems capable of continuous health monitoring. These modern approaches not only improve disease prediction accuracy but also facilitate personalized treatment planning, early intervention, and preventive healthcare, ultimately

III. RESEARCH METHODOLOGY

extraction, model development, performance evaluation, and hybrid model implementation. Each stage is designed to improve prediction accuracy while minimizing noise, redundant information, and model overfitting. The overall workflow ensures that the developed system is reliable, computationally efficient, and suitable for practical healthcare applications.

1. Information Gathering

The hypertension prediction dataset utilized in this study was obtained from the publicly available Kaggle repository. The collection includes 26,083 patient records with a range of clinical and demographic characteristics that influence the risk of hypertension. The primary characteristics include age, gender, type of chest pain, resting blood pressure, cholesterol, fasting blood sugar, electrocardiogram (ECG) results, maximum heart rate attained, exercise-induced angina, ST depression, slope of the ST segment, number of major blood vessels, thalassemia status, and the target class. The objective variable is binary, where 1 indicates the presence of hypertension and 0 indicates its absence. This large dataset provides sufficient information to develop and evaluate machine learning models capable of accurately predicting the risk of hypertension.

2. Data Preparation

Raw healthcare datasets frequently contain incomplete values, inconsistent records, duplicate entries, and irrelevant information, all of which can negatively impact model performance. Consequently, a comprehensive preprocessing stage is completed before model training. Duplicate entries and missing values are first identified and corrected in order to improve the quality of the data. Numerical attributes are normalized to maintain a consistent scale across all variables, while categorical attributes are converted into machine-readable representations. Data cleaning removes outliers and unnecessary noise that could affect classification performance. The processed dataset is then divided into 20% testing data and 80% training data using the train-test split technique. This separation allows for the objective evaluation of model performance while ensuring that the learning algorithms appropriately generalize to new data.

3. Feature Extraction

Feature extraction is an essential process that boosts the efficacy and prediction capacity of machine learning models by selecting the most useful properties from the original dataset. This work uses Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) to minimize feature dimensionality while maintaining significant discriminating information. By dissecting correlated variables into fewer fundamental components, PCA lowers processing complexity and redundancy. LDA further enhances class separability by projecting data into a lower-dimensional space that increases the difference between classes with and without hypertension. These feature extraction techniques improve model accuracy, reduce overfitting, and expedite training.

4. Machine Learning Model Development

Several supervised machine learning algorithms are used to predict hypertension and evaluate the efficacy of their classification. choice Tree generates a hierarchical tree structure that provides clear and intelligible choice criteria by segmenting the data iteratively based on informative qualities. By building a sequence of weak learners, gradient boosting increases classification accuracy by lowering the prediction errors generated by previous models. Extreme Gradient Boosting (XGBoost) enhances the boosting process through regularization, improved tree construction, and parallel processing, enabling faster training and better prediction performance. Random Forest determines the ideal hyperparameters to optimize prediction performance.

5. Hybrid Machine Learning Model

To further improve prediction reliability, a hybrid learning framework is incorporated allowing each algorithm to compensate for the limitations of the others. Random Forest effectively manages noisy and high-dimensional data, Neural Networks capture complex nonlinear relationships among clinical variables, SVM provides strong classification boundaries, and XGBoost enhances overall learning through gradient optimization. The combined framework produces more stable predictions, improves generalization capability, and achieves higher diagnostic accuracy than individual classifiers.

6. Model Evaluation
The effectiveness of the developed models is

7. Overall Workflow
The complete research methodology begins with collecting the hypertension dataset,, after which and robustness. This systematic methodology enables efficient identification of hypertension risk and supports intelligent clinical decision-making through accurate, reliable, and scalable machine learning techniques.

IV. EXISTING SYSTEM

to identify patients who may be at risk of developing high blood pressure. These systems generally utilize clinical parameters have been widely adopted, their prediction performance is often affected by noisy data, redundant features, class imbalance, and limited feature optimization. Many existing models also fail to capture complex nonlinear relationships among multiple health indicators, resulting in reduced prediction accuracy and poor generalization when applied to large and diverse healthcare datasets. Furthermore, several conventional approaches do not incorporate effective dimensionality reduction or feature extraction techniques, leadin. These limitations reduce the reliability of early hypertension diagnosis and highlight the need for more advanced machine learning frameworks capable of providing accurate, robust, and scalable prediction for real-world clinical applications.

V. PROPOSED SYSTEM

by analyzing both clinical and lifestyle-related health parameters. Unlike conventional prediction models that rely on a single classification technique, the proposed approach integrates advancealgorithmsand reliability. Initially, the hypertension dataset is collected and subjected to comprehensive preprocessing, including data cleaning, handling missing values, duplicate removal, normalization, and feature encoding. This preprocessing stage improvesefficient model training.

To enhance the learning capability of the prediction models, dimensionality reduction eliminate redundant attributes while retaining the most informative features. These techniques reduce computational complexity, minimize noise, and improve the discriminative power of the dataset. The optimized data is thenof the developed models.

The proposed framework implements multiple supervised perform hypertension classification. Each algorithm is trained using. Among these models, ensemble learning techniques effectively capture complex nonlinear relationships between patient attributes while reducing overfitting and increasing prediction stability.

Furthermore, the proposed system incorporates by utilizing the robustness of Random Forest, the nonlinear learning capability of Neural Networks, the strong decision boundary of SVM, and the optimization efficiency of XGBoost. The combined framework generates more accurate and consistent predictions than individual models while maintaining high computational efficiency.

hypertension risk prediction. By integrating advanced preprocessing techniques, feature optimization methods, ensemble learning algorithms, and hybrid classification strategies, the framework significantly enhances prediction performance and supports healthcare professionals in early diagnosis, personalized treatment planning, and preventive healthcare management. The methodology maintains the same algorithms and experimental workflow as the base study while presenting the complete approach in a newly written academic format.

VI. SYSTEM ARCHITECTURE

The proposed system architecture presents a comprehensive clinical, demographic, and lifestyle-related information. is obtained from a reliable public repository containing patient attributes such as age, gender, blood pressure, cholesterol level, type, and other health indicators. Since raw healthcare datasets often contain missing values, duplicate records, inconsistent entries, and noisy information, the collected data is subjected to a data preprocessing stage. During preprocessing, missing values are handled, duplicate records are removed, categorical attributes are encoded, numerical features are normalized, and irrelevant

Following preprocessing, the architecture applies feature extraction techniques to improve learning e employed to transform correlated variables into a smaller set of informative components, while Linear Discriminant Analysis (LDA) enhances class separability by selecting the most discriminative features for hypertension classification. These dimensionality reduction techniques remove redundant information, reduce noise, and improve the predictive capability of the machine learning models. The optimized dataset is subsequently provided as input to several supervised. Each classifier independently learns the relationship between patient health parameters and hypertension risk, enabling

To achieve higher prediction accuracy and robustness, the proposed architecture incorporates ensuring a comprehensive assessment of their classification effectiveness. Finally, the optimized hybrid model is deployed as a hypertension prediction system where new patient information is provided as input. Based on the learned patterns, the system predicts whether an individual is likely to be hypertensive or non-hypertensive and generates an accurate risk assessment. This intelligent architecture supports healthcare professionals in early disease detection, timely clinical decision-making, personalized treatment planning, and preventive healthcare management while maintaining high prediction accuracy and reliability.

VII. CONCLUSION

hypertension by utilizing demographic, clinical, and lifestyle-related health information. The proposed methodology combines comprehensive data algorithms to develop a reliable disease prediction system. The implementation of Decision Tree, Gradient Boosting, Random Forest, and Extreme Gradient Boosting (XGBoost) enables a detailed comparative analysis of classification performance, while the incorporation of Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) improves feature quality by eliminating redundant information and reducing data dimensionality. These optimization techniques contribute to enhanced learning efficiency, improved model generalization, and reduced computational complexity.

Experimental evaluation demonstrates that ensemble learning algorithms outperform conventional classification methods in accurately identifying hypertension risk. In particular, Gradient Boosting and XGBoost achieve superior prediction performance, whereas the proposed hybrid model, integrating further enhances classification accuracy, robustness, and reliability. The hybrid framework effectively captures complex relationships among patient health indicators while minimizing overfitting and improving prediction consistency across unseen data.

Overall, the proposed hypertension prediction system provides an efficient and scalable solution for intelligent healthcare applications. By supporting early disease detection and accurate risk assessment, the framework, designing personalized treatment strategies, and promoting preventive healthcare practices. optimized feature engineering demonstrates significant potential for improving patient outcomes and reducing the long-term burden of hypertension-related cardiovascular diseases. Future enhancements healthcare platforms to further improve prediction accuracy, continuous patient monitoring, and intelligent clinical decision support.

REFERENCES

1. "Predicting Hypertension Control Using Machine Learning Techniques," T. Mroz, M. Griffin, R. Cartabuke, C. Laffin, G. Russo-Alvarez, and G. Thomas, PLOS ONE, vol. 19, no. 3, 2024, Art. no. e0299932.
2. "Machine Learning-Based Risk Models for Incident Hypertension Using Population Health Data," F. E. Schjerven, I. Steinsland, F. Lindseth, and E. M. L. Ingeström, Scientific Reports, vol. 14, no. 1, pp. 1–15, 2024.
3. "A Hybrid Machine Learning Framework for Hypertension Risk Prediction," Neural Computing and Applications, vol. 35, no. 20, pp. 14487–14497, 2023; M. Fang, Y. Chen, R. Xue, H. Wang, N. Chakraborty, T. Su, and Y. Dai.
4. "Machine Learning-Based Prediction of Post-Treatment Ambulatory Blood Pressure in Hypertensive Patients," Blood Pressure, vol. 32, no. 1, 2023; H. Hae, S. J. Kang, T. O. Kim, P. H. Lee, S. W. Lee, Y. H. Kim, C. W. Lee, and S. W. Park.
5. "Machine Learning in Hypertension Detection: A Study on World Hypertension Day Data," Journal of Medical Systems, vol. 47, no. 1, 2023, S. Montagna, M. F. Pengo, S. Ferretti, C. Borghi, C. Ferri, G. Grassi, M. L. Muiesan, and G. Parati.
6. "Machine Learning-Based Classification of Hypertension Using Acceleration Photoplethysmography and Clinical Features," S. Abdullah, A. Hafid, M. Lindén, M. Folke, and A. Kristoffersson, Proceedings 328–333 in IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS), 2023.
7. "Hyptn: Predicting Hypertension Using Photoplethysmography Signals and Machine Learning Models," N. Sinha and A. Joshi, IEEE TENCON Conference Proceedings, 2023, pp. 1–6.
8. "Machine Learning Approaches for Predicting Diabetes and Hypertension Using Multiple Clinical Biomarkers," Proceedings of the IEEE Engineering in Medicine and Biology Conference (EMBC), 2023, pp. 1–6, P. Jadhav, V. Selvaraju, S. P. Sathian, and R. Swaminathan.
9. "Machine Learning Techniques for Hypertension Detection and Blood Pressure Estimation: A Comprehensive Review," L. Pecchia, E. Martinez-Ríos, L. Montesinos, and M. Alfaro-Ponce, Biomedical Signal Processing and Control, vol. 68, 2021.
10. "Predicting Hypertension Using Machine Learning: Findings from the Qatar Biobank Study," L. A. AlKaabi, L. S. Ahmed, M. F. Al Attiyah, and M. E. Abdel-Rahman, PLOS ONE, vol. 15, no. 10, 2020.
11. Y. Luo, Y. Li, and Y. Lu, "Deep Learning-Based Hypertension Prediction Using Convolutional Neural Networks," IEEE International Conference on Computer and Communications, 2018, pp. 1123–1128.

12. "Population-Level Machine Learning Models for Hypertension Prediction in South Asia," S. M. S. Islam, A. Talukder, M. A. Awal, M. M. U. Siddiqui, M. M. Ahamad, and B. Ahammed, BMC Public Health, vol. 22, no. 1, pp. 1–14, 2022.
13. "Machine Learning Algorithms for Predicting Hypertension," Y. Sakka, D. Qarashai, and A. Altarawneh, International Journal of Advanced Computer Science and Applications (IJACSA), vol. 14, no. 3, pp. 436–444, 2023.