Proceedings of the 6th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI-2025)
IEEE Xplore Part Number: CFP25US4-ART; ISBN: 979-8-3315-2266-7

Unmasking Hypertension Through a Fusion of Human Lifestyle Data using Machine Learning Algorithms

V. Vishnu Priya
Assistant Professor, Department of Computer Science and Engineering, Kongu Engineering College, Perundurai, Erode, India.
vishnupriya5665@gmail.com

Nisanth G
PG Scholar, Department of Computer Applications Kongu Engineering College Perundurai, Erode, India
nisanthg1010@gmail.com

Nivethaa RP
UG Scholar, Department of Computer Technology-UG Kongu Engineering College Perundurai, Erode, India.
nivethaarp.22bsr@kongu.edu

Srinidhi Vijay
UG Scholar Department of Computer Technology-UG Kongu Engineering College Perundurai, Erode, India
srinidhivijay.22bsr@kongu.edu

2025 6th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI) | 979-8-3315-2266-7/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICMCSI64620.2025.10883119

***Abstract*— Hypertension, also referred to as high blood pressure, is a condition arising from the consistently high blood pressure against artery walls. The volume and output of blood from the heart primarily control blood pressure under stiff or tight arterial conditions. High pulse pressure leads to increased blood pressure, since the pumping of more blood makes the arteries inactive. The World Health Organization has a comprehensive guideline in which an evidence-based pharmacological approach has been recommended for managing adult patients with hypertension, not neglecting to mention lifestyle modification in order to support medicinal treatment. Machine learning may revolutionize the way hypertension treatment is projected for the management of personalized treatment routines for patients based on their characteristics. ML provides great promise in seeking ideal medication and predicting responses thereof against massive sets of data for hypertensive patients. The techniques allow for the determination of the individual risk for hypertension and ensure that treatment protocols are optimal. Compared to methods used to predict hypertension in the past, ML algorithms show a marked improvement in performance.**

***Keywords*—**High Blood Pressure, Artery Walls, Blood Volume, Pulse Pressure, Machine Learning algorithms.

I. INTRODUCTION

Blood pressure leading to hypertension is a worldwide problem in public health, arising from its high prevalence and its permutations of risk for cardiovascular and renal diseases. By supporting the concept of lifestyle modification in preventing clinical hypertension, many studies indicate that individuals at high risk for Williams can be identified at an early stage for necessary early lifestyle interventions to prevent hypertension.

There is some evidence suggesting that the risk of developing hypertension is determined by numerous factors such as advanced age, female sex, elevated body mass index (BMI), family history, premature cardiovascular diseases, sedentary lifestyles, poor food, and high sodium intake.

Nonetheless, the majority of the studies were enhanced by a lack of representativeness of the population, enrolment of small sample sizes, and utilization of a variety of instruments to measure risk factors. A variety of models have been used to successfully identify and stratify patients into risk categories from all risk factors, and then start preventive therapies-for example, the Framingham Risk Score for coronary heart disease statistics and the Clinical Practice Guidelines for Cardiology/American Heart Association Pooled Cohort Equations Risk Calculator for coronary heart disease. However, all these models come with limitations, including non-representative populations, low ethnic diversity, endpoint selectiveness, and low reliability. Thus, the development of population-specific prediction models in South Asian people is warranted.

II. LITERATURE REVIEW

Machine learning (ML) resides as one of the most important tools in hypertension management, providing future approaches for risk assessment and intervention. Support Vector Machines (SVMs), in GUIs, could provide healthcare providers and patients with higher potential development toward decisions to minimize hypertension-related comorbidities, as stated by Essien et al. [1]. Mroz et al.[2] found that SVM application to EHRs in predicting hypertension control takes 75% accuracy. Also, they can improve prediction precision

through the synthetic minority over-sampling technique, as noted by Sakka et al. [3]. Montagna et al.[4] showcased the efficiency of five ML algorithms in detecting hypertension predictors at an approximate accuracy of 90%. The large population surveys from South Asia further investigated the scalability of ML models for self-diagnosis, explained well by Mian Hafeez Ur Rehman et al.[5]. At the same time, the combination of socio-demographic factors and physiological data has been asserted valuable in hypertension prediction, where accuracy rates showed 73% in females and 84% in males [6]. The increased awareness and adherence to hypertension management were, therefore, highlighted by Wilkens et al., in which they achieved 87% study accuracy in their research [7]. For two decades, Volberda et al. propose combining clinical data with physiological signals to achieve 70%-90% accuracy in hypertension prediction [8]. In recent studies, KNN and Light GBM showed 86.8% accuracy in predicting hypertension risk [9]. Recent studies, including Sinha et al. [10], highlight the use of PPG signals and ML for hypertension detection with up to 93% accuracy. Hae et al. further highlighted the utility of ML-based algorithms such as CatBoost to predict post-treatment ambulatory blood pressure with good correlation to ABPM data [11]. Fang et al. reported a hybrid model combining KNN and LightGBM, demonstrating over 86% accuracy for hypertension risk prediction using important blood indicators along with age [12]. Furthermore, Schjerven et al. explored ML models such as SVM and XGBoost. These tools were found to be effective in predicting incident hypertension within an 11-year risk window, especially with features like systolic/diastolic BP and BMI [13]. Future developments in ML, including deep learning and incorporation of diverse data sources, will enhance hypertension detection and management aimed at better health outcomes worldwide.

III. PROPOSED METHODOLOGY

In recent works, for predicting hypertension many machine learning algorithms had been implemented. The proposed methodology includes algorithms like Logistic Regression, Gradient Boosting, XGB, Random Forest. To increase the algorithms' accuracy, certain Feature Extraction techniques like LDA, PCA, Noise and also using cross validation techniques has been employed.

3.1 Data Description

The Hypertension prediction dataset has been gathered using Kaggle, a popular site for machine learning professionals. Users of this portal can search for a dataset as well as published models. The dataset consists of 26083 records in a .csv file. The Hypertension dataset includes age, sex, cp(Chest Pain), trestbps(Resting blood pressure), chol(Cholestrol), fbs(Blood Sugar), restecg(ECG results), thalach(heart rate), exang(angina), old peak(depression), slope, ca(blood vessels), thal, target with class labels zeros and ones. Class zero represents that does not hypertension and one represent the cause hypertension.

3.2 Data Preprocessing

Data preprocessing transforms raw data into meaningful input for machine learning models. Preprocessing techniques includes data reduction, data cleansing, normalization, handling missing values and data integration. After preparing the raw data for further processing, it collects 26,083 records with features such as age,sex,chest pain,cholesterol levels,ecg level and more. Because of an extremely noisy data set, data preprocessing was deemed essential.The train_test_split() breaks data set

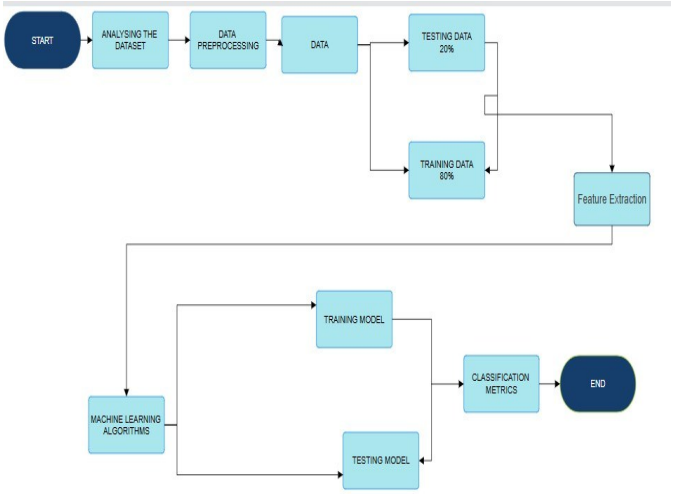

Fig.3.1 Proposed method of ML Investigation

into training set (80%) and testing set (20%), leading to definitions of x_train, x_test, y_train, y_test.x denotes the features, and y is the class label. The model will be trained using X_train and Y_train inputs. Choice enhancement methods would help the model generalize optimally to avoid either underfitting or overfitting of data.

After the training, the model predicts the test set shown in Fig.3.1.This is evaluated by taking classification accuracy score as one way in stating the percentage of correct predictions on the hypertension classification.

3.3 Data Modeling

Decision Tree

Decision tree, a supervised learning technique utilized for resolving regression and classification models, but also frequently used for delivering solutions to classification issues. Decision Tree is a tree-like structured classifier in which the inner nodes indicate the features of dataset, branches indicate decision and each leaf node indicates the

outcome.Decision Tree always starts with a root node and build a tree-like structure it expands on succeeding branches.

Steps to implement decision tree:

STEP 1: Import the Decision Tree classifier in the source code.
STEP 2: Set parameters like n_estimators, learning_rate, minsamples_split, min_samples_leaf and max_depth.
a. The information gain will be calculated using the entropy measure
b. A random_state variable governs the shuffling process.
c. A decision tree can be pre-pruned using max_depth.
STEP 3: Using the fit approach, train the parameters.
STEP 4: Calculate value of X_test.
STEP 5: Finally, Calculate model accuracy.

Gradient Boosting

Gradient Boosting Machine (GBM) is considered one of the popular boosting algorithm in ensemble learning. Gradient Boosting combines multiple weak learners to build a strong model as a final prediction. Gradient Boosting is one of the most powerful ensembling technique that boosts week learner to strong learner. It first constructs a primary model using the training dataset that are available, after which it finds the faults in the base model. In this method a secondary model is constructed after the inaccuracy has been identified, and then a third model is added.

Steps to implement decision tree:

STEP 1: Import the Gradient Boosting classifier in the source code.
STEP 2: Set parameters like random_state, and max_depth.
a. Need to minimize the loss function.
b. Update residuals based on the new ensemble predictions and keep track of each trained weak learner to use them during prediction.
c. Continue the iterative process
STEP 3: Using the fit approach, train the parameters.
STEP 4: Calculate value of X_test.
STEP 5: Finally, Calculate model accuracy.

XGB

XGBoost is an implementation of the gradient boosting framework. The gradient boosting ensemble method constructs predictions in an incremental manner. Each successive model focuses on correcting the mistakes-those that still haven't been predicted by the ensemble of previous models. By correcting the mistakes of previous learners, the gradient boosting model can create accurate predictive outputs from a series of weak deciders, commonly implemented as decision trees.

Steps to implement decision tree:

STEP 1: Import the XGBoosting classifier in the source code.
STEP2: Set parameters like self, learning_rate, n_estimator, max_depth
a.Calculates the residuals for each data point in the training dataset.
STEP 3: By using the fit approach, train the parameters.
STEP 4: Calculate value of X_test.
STEP 5: Finally, calculate model accuracy.

Random Forest

Random Forest, the most popular tree-learning algorithms in an ensemble classifier in Machine Learning and combines several Decision Trees during the training phase. Its construction depends on creating each tree using a random subset of the dataset to measure a random subset of features in each partition. The trees are randomized in order to introduce variability between individual trees, which reduces overfitting and enhances the performance of the overall prediction process**.**

Steps to implement decision tree:

STEP 1: Import the Random Forest classifier in the source code.
STEP 2: Set parameters like self, n_estimator, max_depth.
a.Build decision tree combined with choosen data point.
STEP 3: Using fit approach, train parameters.
STEP 4: Calculate value of X_test.
STEP 5: Finally, Calculate model accuracy.

3.4 **Hybrid Model**

A hybrid model in machine learning is an approach that integrates two or more different types of models or methodologies so as to utilize the strengths of each, giving rise thus to an enhanced, accurate, and more efficient solution shown in Fig.3.2.

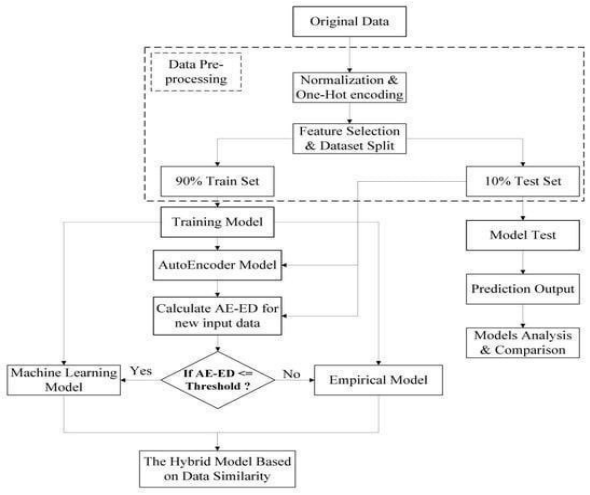

Fig.3.2. Proposed method of Hybrid Model

Random Forest and Neural Network

Hybrid model for predicting hypertension combines random forests and neural networks, leveraging the strengths of both. Randomized methods handle noisy data well, reduce overfitting, and preselect important features, lowering the computational complexity. The neural network then captures complex, non-linear relationships between variables like age, weight, lifestyle, and genetics. This integration reduces variance and bias, improving prediction accuracy and generalization. Additionally, it provides interpretability by highlighting key predictors such as age and BMI, which is valuable for clinicians in understanding model decisions and assessing prediction confidence. Overall, the hybrid approach enhances accuracy, generalization, and interpretability disease prediction.

SVM and XGBoost

A hybrid model combining SVM and XGBoost can effectively predict hypertension by leveraging the strengths of both algorithms. SVM excels at classifying data with clear separations between classes, making it suitable for small to medium datasets. Meanwhile, XGBoost, a gradient-boosting method, handles large datasets and complex feature interactions, while being robust against missing or noisy data. Integrating the two improves classification, reduces overfitting, and enhances generalization. SVM separates hypertensive from non-hypertensive records, while XGBoost optimizes feature interactions and handles noisy data. This combination results in improved prediction accuracy and interpretability, aiding in medical decision-making.

IV. EXPERIMENTAL ANALYSIS

An investigation is carried out among four different machine learning algorithms using certain validation metrics.

4.1 Metrics for Evaluating Classification Models

The calculating metrics like Accuracy, Precision, Recall, F1-Score and ROC-AUC Curve are used for investigation as given in the equation (1),(2),(3),(4) and (5)

Accuracy

Accuracy is used to calculate the number of correct predictions by the total number of predictions.

$$\text{Accuracy} = TP+TN / (TP+TN+FP+FN) \quad \ldots(1)$$

Precision

Precision indicates the proportion of correct predictions of true positives by true positives and false positives.

$$\text{Precision} = TP / (TP+FP) \quad \ldots.(2)$$

Recall

Recall indicates the proportion of correct prediction of true positives by true positives and false negatives.

$$\text{Recall} = TP / (TP + FN) \quad \ldots.(3)$$

F1-Score

F1-Score is calculated by using precision and recall values.

$$\text{F1 score}=\frac{2*(\text{Precision}*\text{Recall})}{(\text{Precision}+\text{Recall})} \quad \ldots(4)$$

ROC-AUC Curve

ROC-AUC is calculated for the efficiency of the model.

$$TPR=TP/(TP+FN)$$
$$FPR=1-\text{Specificity} \quad \ldots (5)$$

4.2 Experimentation

Decision Tree classification studies the contribution of the performance of highly noise-active characteristics to the prediction of a model. The accuracy of 84.36% was found in the trained model, which is impacted by noisy data. Because the technique concentrated on key traits, accuracy increased to 96.57% when it was implemented. The results indicated that feature extraction improved model performance and effectively suppressed noise, which was responsible for a 12.21% increase in accuracy.

Linear Discriminant Analysis as analytical method for Gradient Boosting feature selection. From the outset, the model correctly predicted 86.20% without any duplicates. Upon incorporating LDA, accuracy soared to 98.62% confirming that LDA is capable of reducing features as well as boosting models relatively, gaining a 12.42% accuracy boost.

XGB classification work investigates how the viability of Extreme Gradient Boosting is affected by PCA feature extraction. Prior to PCA cleaning, the model's prime accuracy was 86.54%. All other aspects could be considered extraneous or disruptive, however they primarily fall beyond the bound requirements. Following PCA cleaning, 98.68% accuracy was achieved by the model. This removes the extraneous elements and shrinks the spatial space to the parts that really stand out.

Random Forest assesses how the performance of the Random Forest classifier is affected by the extraction of features using Linear Discriminant Analysis (LDA). With an accuracy of 86.54% prior to LDA implementation, duplicated features may have had an impact on speed. Because LDA improves feature selection by emphasizing on the most discriminative information, accuracy was observed to be 98.68% after employing the algorithm. Since the model's

accuracy rise by a significant 12.14%, the results demonstrate that LDA has a noticeable impact on enhancing the predictive capacity of the model.

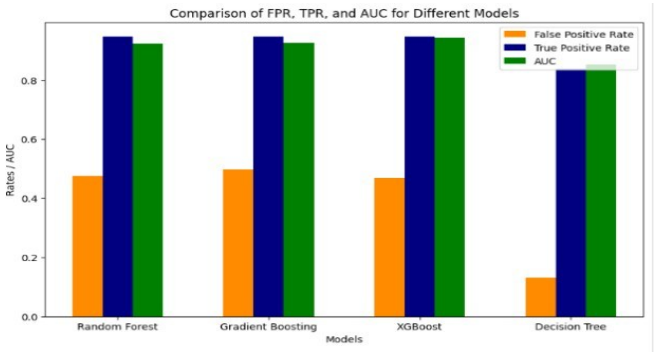

Fig 4.1. Comparison of FPR,TPR and AUC

When building a hybrid model for any given classification problem, as noted in numerous references, the hybrid approach is a sound reason to mix models together to perform much better in any complex decision-making task. An integration of Random Forest, Neural Network, SVM, and Naive Bayes leads to a more nuanced understanding of the data, thus yielding an accuracy of 98.67%. These are best regarded for applications specific to context where accuracy and robustness count, such as medical diagnosis, fraud detection, high-stakes decision making and so forth. To avoid overfitting in the algorithms implement regularization techniques such as L1 and L2 and pruning in decision trees.

Gradient Boosting, XGBoost, Random Forest, and Decision Trees were employed to characterize hypertension management, with Gradient Boosting achieving the highest train and test accuracies of 98.32% and 98.62%, with F1 scoring 98.73%. Therefore, it possessed the basic characteristics of class discrimination. XGBoost followed, having a testing accuracy of 98.68% and F1 Score of 98.79%, indicative of its capability to capture complex patterns in the input data. Random Forest illustrated excellent generalizability with accuracies of 97.74% (training) and 97.76% (testing) and an F1 score of 97.95%. Conversely, Decision Tree appeared to perform lower, showing typical overfitting, exhibiting training and test accuracies of 96.35% and 96.57%, respectively shown in Table 5.1 .

Table 5.1. Performance of Various Machine Learning Algorithms

	Without Feature Extraction				With Feature Extraction			
	XGB	**GBM**	**RF**	**DT**	**XGB**	**GBM**	**RF**	**DT**
Accuracy (%)	86.54	86.20	77.11	84.36	98.68	98.62	97.76	96.57
Precision	83.50	83.02	71.05	84.37	97.60	97.50	98.11	96.57
Recall	93.49	93.49	97.05	84.36	100.00	100.00	97.72	96.57
F1-Score	88.21	87.94	82.03	84.32	98.79	98.73	97.91	96.57
ROC-AUC	0.57	0.48	0.58	0.55	0.57	0.48	0.56	0.50

V. CONCLUSION

Ensemble methods such as Gradient Boosting and XGBoost or Random Forests could prove beneficial in accounting for complex data interactions and could be very relevant to usable application. To improve the detection by implementing ensemble methods and deep learning techniques including CNN or RNN for time series data if it is applicable. Future approaches such as image diagnostics using deep learning may potentially ameliorate early detection of hypertension complications, ensure favourable health conditions for patients, and reduce the worldwide burden of cardiovascular disease. The proposed research introduces a

novel hybrid model combining Random Forest, Neural Networks, SVM and XGBoost for improved accuracy in the risk prediction of hypertension at 98.68%. It also minimize the noise reduce overfitting and focuses on important predictors such as Age and BMI for interpretability.

REFERENCES

[1] Mroz T, Griffin M, Cartabuke R, Laffin L, Russo-Alvarez G, Thomas G, et al. (2024) Predicting hypertension control using machine learning. PLoS ONE 19(3): e0299932.

[2] Yasmin Sakka, Dina Qarashai, Ahmad AltarawnehFaculty of Administrative and Financial Sciences “Predicting Hypertension using Machine Learning”.International Journal of Advanced Computer Science and Applications , Vol. 14,No. 3,2023.

[3] Montagna S, Pengo MF, Ferretti S, Borghi C, Ferri C, Grassi G, Muiesan ML, Parati G. Machine Learning in Hypertension Detection: A Study on World Hypertension Day Data. J Med Syst. 2022 Dec 29;47(1):1. doi: 10.1007/s10916-022-01900-5. PMID: 36580140; PMCID: PMC9800348.

[4] Islam SMS, Talukder A, Awal MA, Siddiqui MMU, Ahamad MM, Ahammed B,. Machine Learning Approaches for Predicting Hypertension and Its Associated Factors Using Population-Level Data From Three South Asian pertention and high blood pressure estimation ” Leandro Pecchia volume68 (2021).

[5] Martinez-Ríos, E., Montesinos, L., Alfaro-Ponce, M. and Pecchia, L., 2021. A review of machine learning in hypertension detection and blood pressure estimation based on clinical and physiological data. *Biomedical Signal Processing and Control*, *68*, p.102813.

[6] AlKaabi LA, Ahmed LS, Al Attiyah MF, Abdel-Rahman ME. Predicting hypertension using machine learning: Findings from Qatar Biobank Study. PLoS One. (2020) PMCID: PMC7567367.

[7] Yue Luo, Yang Li, Yao Lu “The Prediction of hypertension Based CNN” December(2018) DOI:10.1109/CompComm.2018.8780834.

[8] P. Jadhav, V. Selvaraju, S. P. Sathian and R. Swaminathan, "Use of Multiple Fluid Biomarkers for Predicting the Co-occurrence of Diabetes and Hypertension Using Machine Learning Approaches," *2023 45th Annual International Conference of the IEEE* .

[9] S. Abdullah, A. Hafid, M. Lindén, M. Folke and A. Kristoffersson, "Machine Learning-Based Classification of Hypertension using CnD Features from Acceleration Photoplethysmography and Clinical Parameters," *2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS)*,

[10] N. Sinha and A. Joshi, "Hyptn: Predicting Hypertension using PPG signal for Cardiovascular Disease with Machine Learning Models," *TENCON 2023 - 2023 IEEE Region 10 Conference (TENCON)*.

[11] Fang, M., Chen, Y., Xue, R., Wang, H., Chakraborty, N., Su, T. and Dai, Y., 2023. A hybrid machine learning approach for hypertension risk prediction. *Neural Computing and Applications*, *35*(20), pp.14487-14497.

[12] Hae, H., Kang, S.J., Kim, T.O., Lee, P.H., Lee, S.W., Kim, Y.H., Lee, C.W. and Park, S.W., 2023. Machine Learning-Based prediction of Post-Treatment ambulatory blood pressure in patients with hypertension. *Blood Pressure*, *32*(1), p.2209674.

[13] Schjerven, F.E., Ingeström, E.M.L., Steinsland, I. and Lindseth, F., 2024. Development of risk models of incident hypertension using machine learning on the HUNT study data. *Scientific Reports*, *14*(1), p.5609.