A Hybrid Machine Learning Technique based Prediction of Blood Pressure

Smriti Chand, Gagan Preet
M. Tech Scholar, Assistant Professor
Dr. Rishu Bhatia, Professor
Department of Electronics and Communication Engineering,
Ganga Institute of Technology & Management, Kablana Jajjar, India

***Abstract*—Blood pressure is major cause of hypertension and heart health condition which affects a large portion of the global population. Accurate prediction of blood pressure levels plays a crucial role in monitoring and managing this condition effectively. In recent years, machine learning techniques have shown promise in various medical domains, including the prediction of blood pressure. This paper proposes a hybrid machine learning technique for predicting blood pressure levels. The hybrid approach combines the strengths of naive bayes and random forest machine learning algorithms to enhance the accuracy and reliability of blood pressure prediction. Simulation is performed using the python spyder 3.7 software.**

Keywords— Hybrid, Blood pressure, ECG, Heart disease, Pulse rate, Machine learning.

I. Introduction

High blood pressure is a major risk factor for various cardiovascular diseases, including heart attack, stroke, and kidney problems. Accurate prediction of blood pressure levels plays a crucial role in monitoring and managing this condition effectively. By predicting blood pressure, healthcare providers can make informed decisions regarding treatment plans, medication adjustments, and lifestyle modifications for individual patients.

Traditionally, blood pressure prediction has relied on manual measurements taken at clinic visits, which provide only a snapshot of the patient's blood pressure at that specific moment. However, blood pressure is subject to fluctuations throughout the day, influenced by factors such as physical activity, stress, and medication adherence. Therefore, a more comprehensive and continuous approach to blood pressure prediction is needed to capture the dynamic nature of this condition.

In recent years, machine learning techniques have emerged as powerful tools in various medical domains, including disease diagnosis, risk prediction, and treatment optimization. Machine learning leverages the ability of computers to analyze large amounts of data, identify patterns, and make accurate predictions. By applying machine learning algorithms to blood pressure prediction, it is possible to consider multiple factors and their complex interactions, leading to more accurate and personalized predictions.

The objective of this research is to propose a hybrid machine learning technique for predicting blood pressure levels. The hybrid approach combines the strengths of multiple machine learning algorithms, preprocessing techniques, and ensemble learning methods to enhance the accuracy and reliability of the predictions. By integrating different algorithms and leveraging their complementary abilities, the hybrid model aims to capture a wide range of features and patterns in the input data.

The hybrid model consists of two main components of classification model. The classification model determines the blood pressure category, classifying it as normal, prehypertension, or hypertension. By combining regression and classification approaches, the hybrid model provides a comprehensive assessment of blood pressure, enabling both quantitative measurements and categorical classification.

To develop the hybrid model, several steps are involved. Firstly, relevant features are extracted from the input dataset, which may include demographic information (age, gender), medical history (family history of hypertension, presence of other health conditions), and lifestyle factors (smoking, physical activity level). Feature engineering techniques are applied to transform and select the most informative features, enabling the model to focus on the relevant aspects of blood pressure prediction.

The hybrid model incorporates a diverse set of machine learning algorithms to capture different aspects of the data and improve prediction accuracy. Regression algorithms such as linear regression, support vector regression, decision trees, and random forests are employed to model the relationships between the input features and continuous blood pressure values. These algorithms leverage different mathematical principles and assumptions, allowing the hybrid model to capture various patterns and nonlinear relationships present in the data.

Classification algorithms, including logistic regression, random forest, support vector machines, and neural networks, are utilized to classify the blood pressure category accurately. These algorithms are trained to classify the blood pressure levels based on the extracted features and can handle both binary (normal vs. hypertension) and multi-class (normal, prehypertension, hypertension) classification tasks.

II. LITERATURE SURVEY

X. Chen et al.,[1] presented a support vector machine regression model and a random forest regression model for precise blood pressure measurement, both of which help to lessen the impact of individual differences on the model and thus improve the accuracy of blood pressure prediction.

J. Cano et al.[2] purpose of this research is to determine whether machine learning (ML) classifiers can accurately identify hypertension pathophysiology independent of BP data. The objective is to use photoplethysmography (PPG) and electrocardiography (ECG) to differentiate HTS patients from non-HTS recordings and NTS subjects from non-NTS recordings.

D. Chowdhury et al.,[3] purpose of research is to use Machine Learning (ML) techniques to distinguish between ischemia and non-ischemic EST ECGs. EST electrocardiograms were utilised to analyse the hearts of 152 individuals (n=53 females) with a mean age of 5011.92 years. 14 ML classifiers were fed data based on changes in ST morphology recorded pre-load, load, and recovery at J+(40, 60, and 80 ms).

P. Gomathi Shankari et al.,[4] this research, offer a categorization method for stress levels based on HRV frequency domain characteristics. Welch's technique was used to analyse the power spectral density of the ECG data. In addition, the subject's stress levels were analysed using a machine learning model based on the K-Nearest Neighbours (KNN) classifier.

V. S. Arulmurugan et al.,[5] research presented a brand new machine learning-based methodology for predicting systolic and diastolic blood pressure. It uses the 70,000 blood pressure values that are publicly available on the Kaggle website. Different training, validation, and testing percentages were calculated for each of the four used machine learning methods (KNN, logistic regression, decision trees, and random forest) to improve model accuracy.

S. Chen et al.[6] this research was to develop an innovative method for predicting invasive CPP using noninvasive electrocardiogram (ECG) and photoplethysmography (PPG) data by using signal processing and machine learning strategies. The CPP prediction was investigated on three different data sets: (a)ECG alone, (b)PPG only, and (c)both ECG and PPG together.

S. Banerjee et al.,[7] this research, offer a method for estimating blood pressure from Electrocardiograph data using Machine Learning (ML) methods that is amenable to execution on devices with limited computing resources, such as wearables. The suggested approach simply needs non-invasively collected ECG data. The experimental findings demonstrate that the suggested approach outperforms other comparable methods reported in the literature.

C.-T. Yen, et al.,[8] presented a sufficiently accurate cuffless blood pressure (BP) estimate technique based on photoplethysmography (PPG) and electrocardiography (ECG) data, given that existing cuffless BP monitoring methods have adequate overall accuracy. In this investigation, estimated HR, diastolic BP, and systolic BP using single-channel PPG and electrocardiogram (ECG) data. To create an accurate deep learning model for predicting BP and HR, combined a multi-scale convolution network with a long short-term memory (LSTM) network.

I. Kuzmanov et al.,[9] appears reasonable to use both signals for blood pressure categorization since they both reveal unique characteristics of the cardiovascular system. In large-scale emergencies like natural disasters, when many people are wounded, it is crucial to quickly assess their blood pressure as part of the triage procedure in order to monitor their hemostability.

T. Dave & coauthors[10] approach suggested herein utilises a wireless hardware device to capture ECG and PPG data, from which time domain characteristics are extracted. A lightweight model for Blood Pressure estimation is developed using the machine learning technique of Support Vector Regression. Wireless signals from 87 participants were acquired using a hardware device to test the planned concept.

Durga, B. Siva Praveen, et al., [11] the electrocardiogram (ECG) is a cuff-based technique for estimating or classifying blood pressure. Photoplethysmography (PPG) signals are being investigated as a cuff-free, non-invasive alternative. Pulse oximetry (PPG) is an optical, noninvasive technique for measuring pulse-to-pulse variations in blood volume. The PPG signal may also be interpreted as a reflection of the heart's mechanical activity.

A. I. Hossain, et al., [12] this research uses five machine learning algorithms to predict heart illness using a dataset of 1190 records obtained from the UCI repository: the Support Vector Machine, Logistic Regression, K-nearest Neighbour, Naive Bayes, and Ensemble Voting Classifier.

III. METHODOLOGY

The following flowchart explains the suggested methodology:-

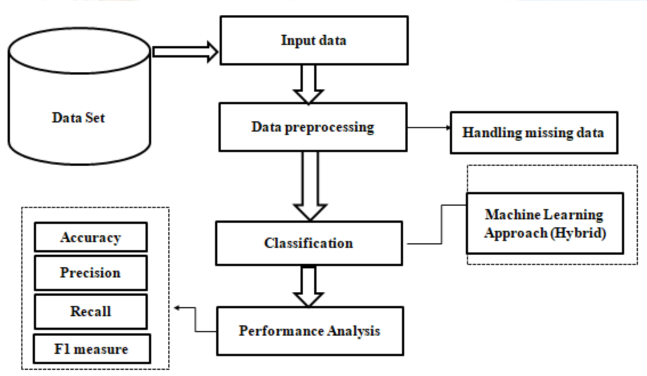

Figure 1: Flow Chart

Steps-

- Firstly, In the first place, complete the dataset [13] derived from the a publicly accessible, huge dataset source.
- The data has been preprocessed, and the missing dataset is being sent over right now. Get rid of the blank spot by replacing it with a 1 or 0.
- Next, use a classification approach that takes use of the advantages of both traditional machine learning and the more recent hybrid approach.
- Precision, recall, F-measure, accuracy, and error rate are some of the performance metrics you should now examine and compute.

The suggested research technique relies on the following sub modules:

Data Selection and Loading

- The data selections are the process of selecting the dataset and load this dataset into the python environment.

Data Pre-processing

- Data selection is the process of choosing a dataset and importing it into a Python environment.
- The First Steps in Processing Data
- Unwanted information is filtered out of a dataset at the pre-processing stage.

Splitting Dataset into Train and Test Data

- Data splitting is the process of dividing a dataset into two halves, often for use in a cross-validator.
- The data is split in half, with one half used to create a prediction model and the other half for testing how well that model performed.

Feature Extraction-

Data independence may be standardised via the use of feature extraction. It's often done in the pre-processing phase of data analysis and is also known as normalisation.

Classification- When it came to making the distinctions, a mixture of random forest and a naive bayes classifier was used.

The hybrid technique based on Random Forest and Naive Bayes for the prediction of blood pressure combines the strengths of both algorithms to improve the accuracy and reliability of blood pressure predictions. Random Forest is an ensemble learning method that constructs a multitude of decision trees and aggregates their predictions, while Naive Bayes is a probabilistic classifier based on Bayes' theorem. By integrating these two algorithms, we can benefit from the robustness of Random Forest and the probabilistic nature of Naive Bayes in the prediction of blood pressure.

The hybrid technique follows a two-step process. In the first step, a Random Forest model is trained using a dataset that includes relevant features such as demographic information, medical history, and lifestyle factors. Random Forest constructs multiple decision trees based on bootstrapped samples of the data and random subsets of features. Each decision tree is trained independently, and the final prediction is obtained by aggregating the predictions of all the trees, either through voting or averaging. Random Forest is known for its ability to handle high-dimensional data, capture non-linear relationships, and handle noisy or missing data effectively.

In the second step, Naive Bayes is applied as a post-processing step to refine the predictions generated by the Random Forest model. Naive Bayes calculates the probabilities of each class label (e.g., normal, prehypertension, hypertension) based on the observed features. It assumes that the features are independent of each other, although this assumption may not hold in practice. Nevertheless, Naive Bayes can still provide valuable insights by considering the conditional probabilities of the features given each class label. The final prediction is obtained by selecting the class label with the highest probability.

The integration of Random Forest and Naive Bayes in the hybrid technique combines the strengths of both algorithms. Random Forest captures complex relationships and interactions between features, while Naive Bayes introduces a probabilistic component that can further refine the predictions based on the conditional probabilities. This hybrid approach aims to improve the accuracy and robustness of blood pressure predictions by leveraging the complementary abilities of the two algorithms.

IV. SIMULATION RESULT

The implementation of the proposed algorithm is done over python spyder 3.7. The sklearn, numpy, pandas, matplotlib, pyplot, seaborn, os library helps us to use the functions available in spyder environment for various methods like decision tree, random forest, naive bayes etc.

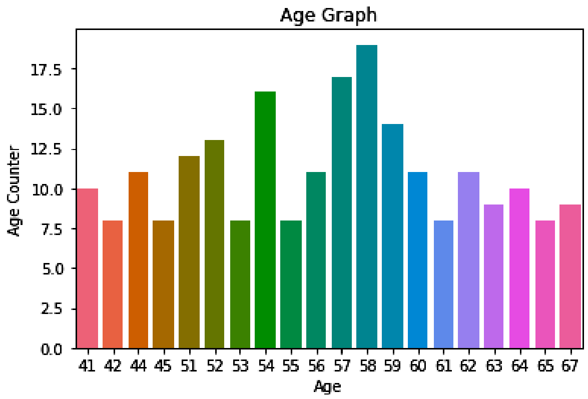

Figure 3: Graph of data feature- Age

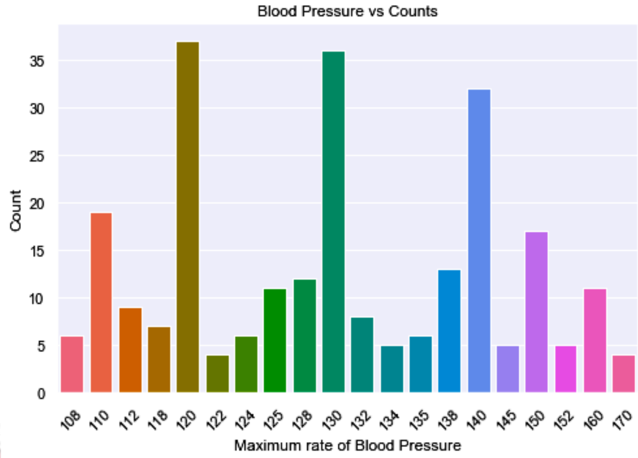

Figure 4: Blood pressure counts

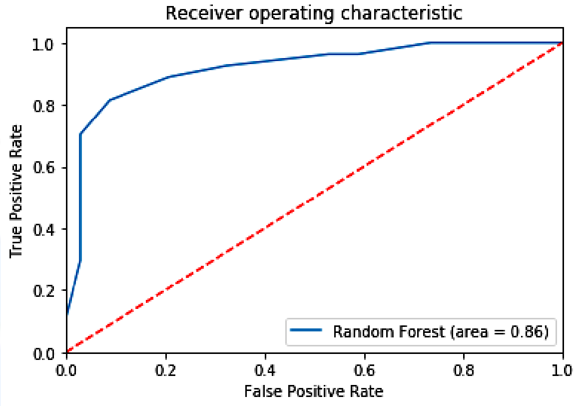

Figure 5: ROC result graph of Random forest

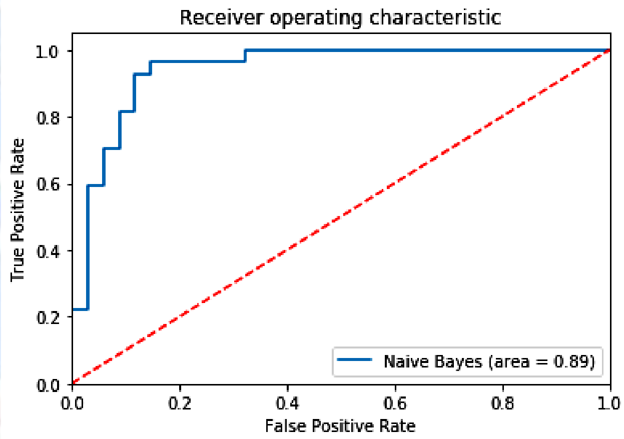

Figure 6: ROC result graph of Naïve Bayes

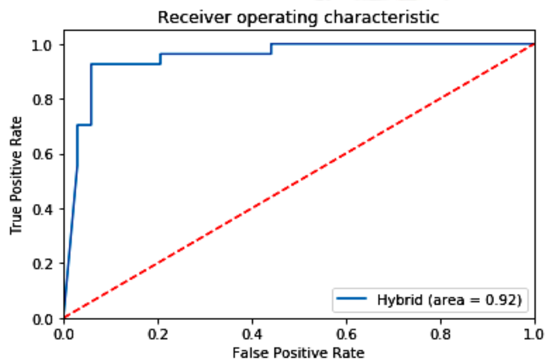

Figure 7: ROC result graph of proposed method (Hybrid)

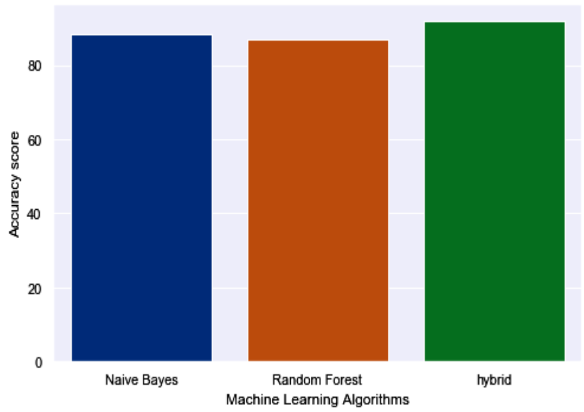

Figure 8: Result comparison of applied methods

From all results graph figures and receiver operating characterises curve, it is clear the proposed hybrid algorithm gives significant improved results than individual random forest and naïve bayes methods.

Table 1: Simulation Result of Random Forest

Sr. No.	Parameters	Value
1	Accuracy	86
2	Classification error	14
3	Precision	86
4	Recall	91
5	F-measure	89

Table 2: Simulation Result of Naive Bayes

Sr. No.	Parameters	Value
1	Accuracy	89
2	Classification error	11
3	Precision	91
4	Recall	88
5	F-measure	90

Table 3: Simulation Result of Proposed Hybrid method

Sr. No.	Parameters	Value
1	Accuracy	92
2	Classification error	8
3	Precision	94
4	Recall	91
5	F-measure	93

Table 1, 2 and 3 are showing simulation results when applying random forest, naïve bayes and proposed hybrid algorithm.

V. CONCLUSION

A hybrid machine learning technique combining Random Forest and Naive Bayes algorithms shows promise for predicting blood pressure. The combination of Random Forest's ability to handle complex data sets, capture nonlinear relationships, and provide feature importance, along with Naive Bayes' efficiency in dealing with high-dimensional data and probabilistic predictions, can result in improved accuracy and robustness. By leveraging the strengths of both algorithms, the hybrid approach can enhance the predictive power of blood pressure models. This can be especially valuable in the field of healthcare, where accurate predictions of blood pressure can aid in diagnosing and managing various conditions.

REFERENCES

1. X. Chen, S. Yu, Y. Zhang, F. Chu and B. Sun, "Machine Learning Method for Continuous Noninvasive Blood Pressure Detection Based on Random Forest," in IEEE Access, vol. 9, pp. 34112-34118, 2021, doi: 10.1109/ACCESS.2021.3062033.
2. J. Cano, V. Bertomeu-González, L. Fácia, J. Moreno-Arribas, R. Alcaraz and J. J. Rieta, "ECG and PPG-Based Hypertension Screening Under Non-Hypertensive Blood Pressure Recordings," 2022 Computing in Cardiology (CinC), Tampere, Finland, 2022, pp. 1-4, doi: 10.22489/CinC.2022.093.
3. D. Chowdhury, B. C. Neelapu, K. Pal and J. Sivaraman, "Machine Learning-based Classification of Ischemic and Non-Ischemic Exercise Stress Test ECG," 2022 Computing in Cardiology (CinC), Tampere, Finland, 2022, pp. 1-4, doi: 10.22489/CinC.2022.276.

4. P. Gomathi Shankari, D. A. Hrithik, D. Ravikumar, D. Chaitanya, S. Pushpa Mala and S. Kokila, "Heart Rate Variability and Machine Learning for Stress Analysis," 2022 4th International Conference on Circuits, Control, Communication and Computing (I4C), Bangalore, India, 2022, pp. 73-78, doi: 10.1109/I4C57141.2022.10057650.
5. V. S. Arulmurugan, C. Rajeswari, A. Bhardwaj, G. Shobana, A. Dubey and S. Mujoo, "Machine Learning Based Systolic and Diastolic Pressure Prediction Model," 2022 International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), Trichy, India, 2022, pp. 461-466, doi: 10.1109/ICAISS55157.2022.10010729.
6. S. Chen et al., "Predicting Invasive Coronary Perfusion Pressure Using Noninvasive Electrocardiogram and Photoplethysmography Based on Machine Learning for Cardiopulmonary Resuscitation," 2022 15th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Beijing, China, 2022, pp. 1-6, doi: 10.1109/CISP-BMEI56279.2022.9980348.
7. S. Banerjee, B. Kumar, A. P. James and J. N. Tripathi, "Blood Pressure Estimation from ECG Data Using XGBoost and ANN for Wearable Devices," 2022 29th IEEE International Conference on Electronics, Circuits and Systems (ICECS), Glasgow, United Kingdom, 2022, pp. 1-4, doi: 10.1109/ICECS202256217.2022.9970924.
8. C. -T. Yen, S. -N. Chang and C. -H. Liao, "Estimation of Beat-by-Beat Blood Pressure and Heart Rate From ECG and PPG Using a Fine-Tuned Deep CNN Model," in IEEE Access, vol. 10, pp. 85459-85469, 2022, doi: 10.1109/ACCESS.2022.3195857.
9. I. Kuzmanov, A. M. Bogdanova, M. Kostoska and N. Ackovska, "Fast Cuffless Blood Pressure Classification with ECG and PPG signals using CNN-LSTM Models in Emergency Medicine," 2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO), Opatija, Croatia, 2022, pp. 362-367, doi: 10.23919/MIPRO55190.2022.9803582.
10. T. Dave, U. Pandya and M. Joshi, "Cuff-less Blood Pressure measurement from Wireless ECG and PPG signals," 2021 IEEE International Symposium on Smart Electronic Systems (iSES), Jaipur, India, 2021, pp. 33-37, doi: 10.1109/iSES52644.2021.00020.
11. B. Durga Siva Praveen, D. V. N. Sandeep, I. V. V. Raghavendra, M. Yuvaraj and S. Sarath, "Non-invasive Machine Learning approach for classifying Blood Pressure using PPG Signals in COVID situation," 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2021, pp. 1-7, doi: 10.1109/ICCCNT51525.2021.9579535.
12. A. I. Hossain, S. Sikder, A. Das and A. Dey, "Applying Machine Learning Classifiers on ECG Dataset for Predicting Heart Disease," 2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI), Rajshahi, Bangladesh, 2021, pp. 1-6, doi: 10.1109/ACMI53878.2021.9528169.