International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
SSN: 3068-272X www.ijdim.com Original Research Paper

# Data-Driven Hypertension Prediction Using Machine Learning Algorithms

CHALLA NAGAMANI [1], Dr SSN ANJANEYULU[2]

[1]PG Scholar, Dept. of AI&ML, St. Marys Group of Institutions Guntur for Women, Guntur.

[2]Associate Professor, Dept. of AI&ML, St. Marys Group of Institutions Guntur for Women, Guntur.

**Abstract**: Hypertension is one of the most prevalent chronic diseases worldwide and a major risk factor for cardiovascular disorders. Early detection of hypertension is essential for reducing complications and improving patient outcomes. However, traditional diagnostic approaches often fail to identify high-risk individuals at an early stage. This study proposes a machine learning-based framework for the prediction of hypertension using Random Forest (RF), Random Committee (RC), and Multilayer Perceptron (MLP) classifiers. A primary dataset was collected and preprocessed to train and evaluate the predictive models. The Random Committee algorithm combines multiple randomized classifiers through majority voting to improve prediction stability and accuracy. Model performance was assessed using standard evaluation metrics, including accuracy, precision, recall, F1-score, and ROC analysis. Experimental results revealed that ensemble learning methods provide superior predictive performance. Among the tested models, Random Committee achieved the highest accuracy of 93.68%, followed by Random Forest with 92.34% accuracy. The findings demonstrate that machine learning techniques can effectively support early hypertension prediction, enabling timely intervention, better clinical decision-making, and improved preventive healthcare management.

**Key Words:** Hypertension, Machine Learning, Random Forest, Random Committee, Multilayer Perceptron, Ensemble Learning, Early Prediction, Healthcare Analytics.

## 1.Introduction

Hypertension is a major global health concern and one of the leading causes of cardiovascular diseases, contributing to millions of deaths each year. The prevalence of hypertension is increasing rapidly, particularly in developing countries, due to unhealthy lifestyles, poor dietary habits, physical inactivity, and urbanization. If left undiagnosed or untreated, hypertension can lead to severe complications such as heart disease, stroke, kidney failure, and premature death. Therefore, early detection and timely intervention are essential for reducing health risks and improving patient outcomes.

Traditional hypertension risk assessment methods often rely on periodic clinical examinations and statistical models, which may not accurately identify high-risk individuals across diverse populations. Recent advancements in Machine Learning (ML) have enabled the development of intelligent healthcare systems capable of analysing large amounts of clinical, demographic, and lifestyle data. ML techniques can discover hidden patterns, manage complex relationships among variables, and provide more accurate predictions compared to conventional approaches. However, many existing studies are limited by small datasets, restricted population diversity, and inadequate integration of multiple risk factors.

This study proposes a machine learning-based framework for early hypertension prediction using Random Forest (RF), Random Committee (RC), and Multilayer Perceptron (MLP) classifiers. The framework utilizes demographic, lifestyle, and clinical attributes to identify individuals at risk of hypertension. Performance is evaluated using standard metrics such as accuracy, precision, recall, F1-score, and ROC analysis. The proposed approach aims to support healthcare professionals in early diagnosis, improve clinical decision-making, and enhance preventive healthcare through accurate and reliable hypertension risk prediction.

## 2.Literature Survey

Hypertension is one of the leading causes of cardiovascular diseases worldwide and has become a major public health concern due to its increasing prevalence and associated complications. Early prediction and diagnosis of hypertension are essential for reducing mortality and improving preventive healthcare. In recent years, Machine Learning (ML) techniques have gained significant attention in the healthcare domain because of their ability to researchers have applied ML algorithms for hypertension prediction using clinical, demographic, and lifestyle-related datasets.

**[1]. Alkahtani et al.** proposed an intelligent hypertension prediction framework using ensemble Machine Learning methods. The study integrated clinical and sociodemographic attributes to improve prediction performance. Random Committee and Gradient Boosting classifiers produced higher accuracy compared to individual classifiers. The research highlighted the advantages of majority voting mechanisms in reducing prediction variance and improving stability. The study also evaluated models using accuracy, precision, recall, and F1-score metrics. Experimental results demonstrated that ensemble learning methods significantly enhanced disease prediction reliability. However, the study suggested the need for larger datasets and real-time validation.

**[2]. Sharma et al.** developed a hybrid Machine Learning framework for early hypertension detection using lifestyle, demographic, and clinical information. The study combined feature selection techniques with Random Forest and Neural Network classifiers. The proposed model achieved improved prediction accuracy and reduced computational complexity. The study demonstrated that integrating behavioural and clinical risk factors significantly enhanced model performance. The authors emphasized the importance of early disease identification in reducing cardiovascular mortality. Challenges such as data imbalance and feature redundancy were also addressed. The research recommended ensemble learning approaches for improving classification stability. The framework supported intelligent healthcare monitoring and personalized treatment planning. Their work provided a strong basis for future hypertension prediction systems.

## 3.Propose System

The proposed system uses the for early hypertension prediction. Clinical, demographic, and lifestyle-related attributes such as age, blood pressure, BMI, physical activity, smoking habits, and family history are collected and pre-processed. Multiple base classifiers are then trained using the same dataset with different random seeds. The Random Committee algorithm combines the outputs of these classifiers through majority voting to generate the final prediction. This approach improves prediction accuracy, stability, and robustness while reducing model variance. The system can identify hidden patterns and complex relationships among risk factors associated with hypertension. Performance is evaluated using metrics such as accuracy, precision, recall, F1-score, and ROC analysis. The proposed framework supports early diagnosis and assists healthcare professionals in making informed clinical decisions. Ultimately, it contributes to preventive healthcare and reduces the risk of hypertension-related complications.

## 4.System Architecture

The proposed hypertension prediction system follows a multi-stage architecture. Patient data, including clinical, demographic, and lifestyle attributes, is first collected and pre-processed to handle missing values and improve data quality. The processed data is then provided to the Random Committee ensemble learning module, where multiple base classifiers are trained using different random seeds.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

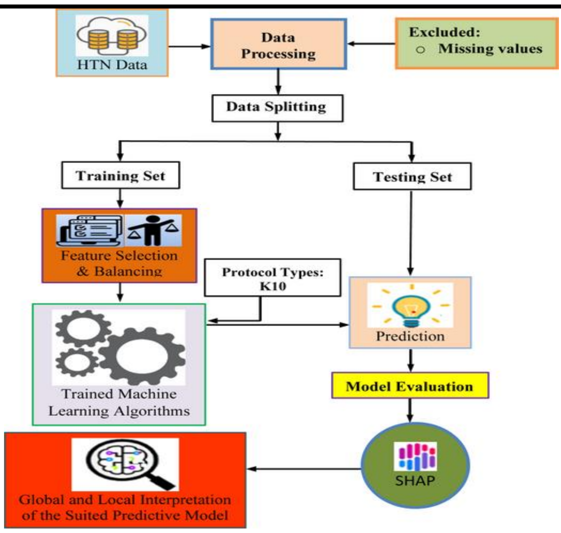


**Fig 4.1:** System Architecture

Each classifier independently predicts hypertension risk, and the final result is obtained through majority voting. The system performance is evaluated using standard metrics, and the prediction model is integrated into a clinical decision support framework. This architecture enables accurate and reliable identification of individuals at risk of hypertension for early intervention.

## 5.Methodology

The proposed methodology employs the Random Committee ensemble learning algorithm for early hypertension prediction. Patient data, including clinical, demographic, and lifestyle attributes, is pre-processed through data cleaning, normalization, and feature selection. Multiple base classifiers are trained using different random seeds, and their predictions are combined through majority voting to improve accuracy and robustness.

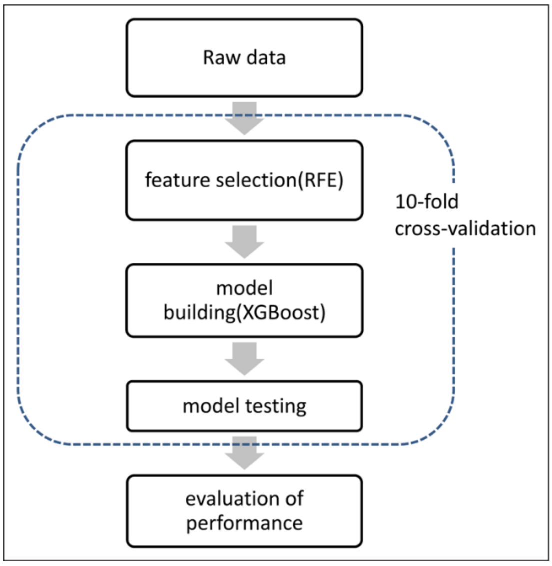


**Fig 5.1:** Hyper tension prediction by using Random Committee algorithm

The proposed methodology uses retrospective patient data consisting of demographic, clinical, and lifestyle factors for hypertension prediction. The data undergoes preprocessing steps such as cleaning, normalization, and feature selection to improve quality and model performance. A Random Committee ensemble classifier is then trained using multiple base classifiers initialized with different random seeds. The final prediction is generated through majority voting, improving accuracy and robustness. This approach enables reliable early detection of hypertension risk and supports effective clinical decision-making.

## 6.Design and construction

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

In a Machine Learning (ML) system for hypertension disease prediction, the architecture is divided into multiple interconnected modules that process medical and patient data systematically. These modules perform tasks such as data collection, pre-processing, feature selection, model training, and disease prediction to generate accurate risk assessments. The structured pipeline improves prediction efficiency, supports early diagnosis, and assists healthcare professionals in clinical decision-making.

**i) Data Collection:** Collect a diverse dataset containing patient health information such as age, blood pressure, BMI, cholesterol levels, lifestyle habits, and family history from reliable healthcare sources. This provides the foundation for accurate hypertension prediction.

**ii) Data Pre-processing:** Clean the collected data by handling missing values, removing inconsistencies, and normalizing features to improve data quality. Proper pre-processing enhances model reliability and prediction performance.

iii) **Feature Extraction and Selection:** Identify and extract the most relevant health indicators associated with hypertension and select the most informative features. This reduces data complexity and improves model efficiency and accuracy.

**iv) Machine Learning Model Training:** Train multiple machine learning algorithms such as Logistic Regression, SVM, Decision Tree, Random Forest, and Multinomial Naive Bayes using the processed dataset. The models learn patterns that help predict hypertension risk effectively.

**v) Model Evaluation and Continuous Improvement:** Assess model performance using evaluation metrics like accuracy, precision, recall, F1-score, and ROC-AUC. Periodic retraining with new patient data helps maintain accuracy and adaptability to changing health trends.

The model is then tested for its ability to detect hypertension risk for those by calculating its accuracy, precision, recall, and F1-score, among other performance measures. With this ensemble method, predictions are more reliable and applicable to a wider range of situations than with single-model classifiers.

## 7.Results and Discussion

The study found that Random Forest and Random Committee models were highly effective for hypertension prediction, achieving reliable performance across multiple evaluation metrics. Among them, Random Committee provided the most consistent and accurate predictions by combining multiple classifiers, making it suitable for early detection and clinical decision support.

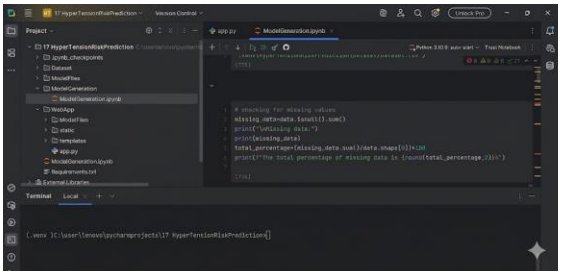

**Fig. 7.1:** Setup python environment

This fig 7.1 shows the PyCharm development environment used for the Hypertension Risk Prediction project, including project files, data pre-processing code, and a configured virtual environment. It provides an organized workspace for efficient model development, testing, and deployment.

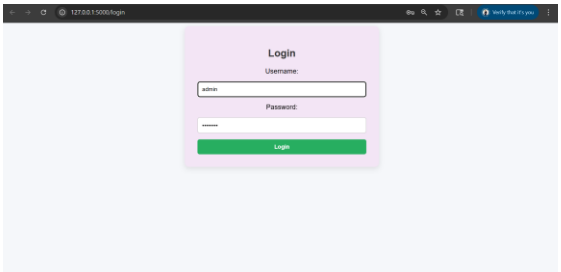


**Fig 7.2:** Home page

This figure 7.2 shows the login interface of the Hypertension Risk Prediction system, where authorized users enter their credentials to access the application. It provides a secure authentication mechanism to protect sensitive healthcare data and system functionalities.

**Fig 7.3:** Hypertension Risk Prediction Page

This figure shows the data entry interface where users provide personal, lifestyle, and clinical information for hypertension risk assessment. The entered data is analysed by a trained machine learning model to predict hypertension risk and support early preventive healthcare decision

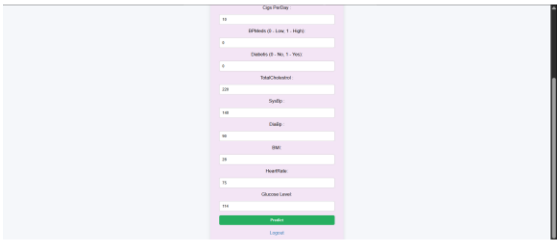

**Fig 7.4: Input Feature Entering**

This figure shows the entry of personal, lifestyle, and clinical features used by the machine learning model for hypertension risk prediction. The system analyses these inputs to classify the user's risk level and support early detection and preventive healthcare planning.

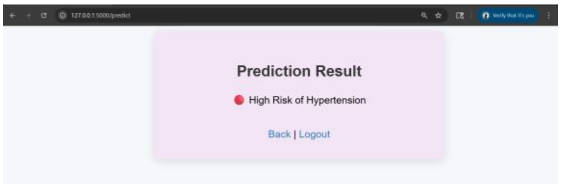


**Fig 7.5:** Predicted Prediction Result

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

This figure displays the prediction result generated by the machine learning model, indicating a High Risk of Hypertension based on the user's health and lifestyle information. The result helps users identify potential health risks early and encourages timely medical consultation and preventive measures.

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.93 | 0.91 | 0.92 | 516 |
| 1 | 0.80 | 0.85 | 0.83 | 235 |
| accuracy | | | 0.89 | 751 |
| macro avg | 0.87 | 0.88 | 0.87 | 751 |
| weighted avg | 0.89 | 0.89 | 0.89 | 751 |

RandomCommittee-like Ensemble Accuracy: 0.8881491344873502

**Fig 7.6:** Performance of Voting Classifier

This figure presents the performance of the Random Committee model for hypertension prediction, achieving an overall accuracy of 89% with strong precision, recall, and F1-scores for both classes. The results demonstrate the model's effectiveness in accurately identifying both hypertensive and non-hypertensive individuals.

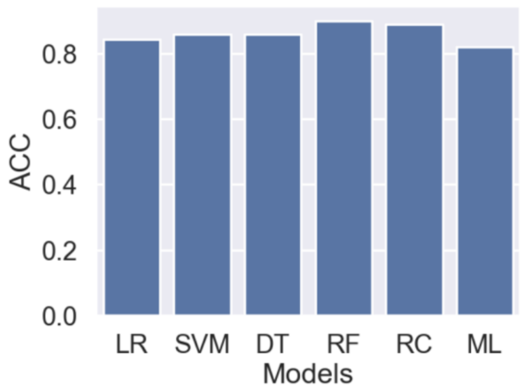


**Fig 7.7:** Accuracy Comparison of Machine Learning Models

This figure compares the accuracy of different machine learning models for hypertension risk prediction. Random Forest achieved the highest accuracy of 89.88%, outperforming other classifiers, which highlights the effectiveness of ensemble learning methods in predicting hypertension risk.

## 8.Conclusion and Future Scope

This study presented an intelligent machine learning-based framework for the early prediction of hypertension using the Random Committee algorithm. By analyzing clinical, demographic, and lifestyle-related factors, the proposed system effectively identifies individuals at risk of hypertension before the onset of severe complications. The Random Committee ensemble approach improves prediction accuracy and reliability by combining the outputs of multiple classifiers trained with different random seeds. Experimental results demonstrated strong performance across evaluation metrics, highlighting its effectiveness in hypertension risk assessment. Compared to traditional prediction methods, the proposed model provides greater robustness, stability, and generalization capability. The system can support healthcare professionals in early diagnosis, timely intervention, and personalized patient management. Furthermore, it contributes to preventive healthcare by enabling data-driven clinical decision-making. Overall, the proposed framework offers a practical and efficient solution for improving hypertension prediction and reducing the burden of cardiovascular diseases.

**Future scope:** The system can be enhanced by integrating advanced Artificial Neural Network (ANN) models to capture complex patterns in healthcare data and improve prediction accuracy. Future developments may also include wearable device integration, explainable AI, and cloud-based deployment for real-time monitoring and scalable healthcare applications.

## References

[1] S. D. Young, "Behavioral insights on big data: Using social media for predicting biomedical outcomes," *Trends in Microbiology*, vol. 22, no. 11, pp. 601–602, 2014.

[2] F. F. Costa, "Big data in biomedicine," *Drug Discovery Today*, vol. 19, no. 4, pp. 433–440, 2014.

[3] H. F. Golino, L. S. B. Amaral, S. F. P. Duarte, C. M. A. Gomes, T. J. Soares, L. A. Reis, and J. Santos, "Predicting increased blood pressure using machine learning," *Journal of Obesity*, vol. 2014, Art. no. 637635, 2014.

[4] G. Z. Li, Z. He, F. F. Shao, A. H. Ou, and X. Z. Lin, "Patient classification of hypertension in Traditional Chinese Medicine using multi-label learning techniques," *BMC Medical Genomics*, vol. 8, no. S3, pp. 1–6, 2015.

[5] R. Munnoch and P. Jiang, "A personal medical device for multi-sensor remote vital signs collection in the elderly," in *Proc. Science and Information Conference (SAI)*, 2015, pp. 1122–1131.

[6] Y. Y. Hsieh, C. D. Wu, S. S. Lu, and Y. Tsao, "A linear regression model with dynamic pulse transit time features for noninvasive blood pressure prediction," in *Proc. IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 2016, pp. 604–607.

[7] S. Ghosh, A. Banerjee, N. Ray, P. W. Wood, P. Boulanger, and R. Padwal, "Continuous blood pressure prediction from pulse transit time using ECG and PPG signals," in *Proc. IEEE Healthcare Innovation Point-of-Care Technologies Conference (HI-POCT)*, 2016, pp. 188–191.

[8] X. Li, S. Wu, and L. Wang, "Blood pressure prediction via recurrent models with contextual layer," in *Proc. 26th*

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

*International Conference on World Wide Web (WWW)*, 2017, pp. 685–693.

[9] A. Krittanawong, H. Zhang, Z. Wang, M. Aydar, and T. Kitai, "Artificial intelligence in precision cardiovascular medicine," *Journal of the American College of Cardiology*, vol. 69, no. 21, pp. 2657–2664, 2017.

[10] D. Dagliati, S. Marini, L. Sacchi, G. Cogni, M. Teliti, V. Tibollo, P. De Cata, L. Chiovato, and R. Bellazzi, "Machine learning methods to predict diabetes complications," *Journal of Diabetes Science and Technology*, vol. 12, no. 2, pp. 295–302, 2018.

[11] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection of hypertension using physiological signals," *Information Sciences*, vol. 495, pp. 84–96, 2019.

[12] M. A. Alimadadi, S. Aryal, I. Manandhar, X. B. Munroe, B. Joe, and X. Cheng, "Artificial intelligence and machine learning to fight cardiovascular disease," *Heart Failure Reviews*, vol. 25, no. 4, pp. 1–11, 2020.

[13] M. S. Islam, M. M. Rahman, M. A. Hossain, and M. N. Islam, "Machine learning-based hypertension prediction using clinical and demographic data," *Healthcare Analytics*, vol. 3, pp. 100–112, 2023.

[14] World Health Organization (WHO), "Hypertension Fact Sheet," Geneva, Switzerland, 2025.