Implementation of Hybrid Machine Learning (SVM, Decision Tree, Random Forest) Algorithms for Early Hypertension Prediction

Sk. Himam Basha[1], P. Satya Sri Lakshmi[2]
Associate Professor[1], MCA Scholar[2]
Dept. of Master of Computer Applications,
QIS College Of Engineering & Technology(Autonomous)
Ongole, AP, India

Abstract

Hypertension, commonly known as high blood pressure, is a major risk factor for cardiovascular diseases and premature mortality worldwide. Early detection and prevention are critical in reducing its health impact. This study explores the application of machine learning (ML) techniques to predict the likelihood of hypertension in individuals using clinical and demographic data. A variety of supervised learning algorithms, including Logistic Regression, Random Forest, Support Vector Machines, and Gradient Boosting, were evaluated for their predictive performance. The dataset was preprocessed through feature selection, normalization, and handling of missing values to improve model accuracy. Performance metrics such as accuracy, precision, recall, F1-score, and AUC-ROC were used to assess the models. The results demonstrate that ML models can effectively identify individuals at high risk of hypertension, offering a valuable tool for early intervention and personalized healthcare. This approach underscores the potential of artificial intelligence in supporting public health efforts and enhancing clinical decision-making.

Introduction

Hypertension, commonly known as high blood pressure, is a major public health concern globally. It is often referred to as a "silent killer" because it typically has no

symptoms until it has caused significant damage to the heart, arteries, or other organs. Early detection and management are critical to preventing complications such as heart disease, stroke, and kidney failure. Traditional methods of diagnosing hypertension rely heavily on regular medical check-ups and patient-reported symptoms, which may not always be timely or accurate.

With the advancement of technology and the availability of large-scale health data, machine learning (ML) offers a promising approach to enhance the early prediction and diagnosis of hypertension. Machine learning algorithms can identify complex patterns in data, learn from historical records, and make accurate predictions based on input features such as age, weight, lifestyle factors, and medical history.

This study explores the application of machine learning techniques for the prediction of hypertension. By leveraging data-driven approaches, we aim to develop predictive models that can assist healthcare professionals in identifying individuals at risk, thereby enabling timely interventions

and personalized healthcare strategies. The integration of machine learning into hypertension prediction represents a significant step toward more proactive, efficient, and intelligent healthcare systems.

Literature Survey

1. Importance of Early Detection of Hypertension

- Hypertension is a major risk factor for cardiovascular diseases, stroke, and kidney failure.
- Traditional diagnosis relies on periodic monitoring and physical check-ups, which may delay early intervention.
- Literature emphasizes the need for predictive tools to enable timely identification and management.

2. Common Machine Learning Techniques Used

- Studies frequently use supervised learning models like **Logistic Regression, Decision Trees, Random Forest, SVM, KNN, and Neural Networks**.
- Ensemble techniques such as **XGBoost and LightGBM** have shown improved performance.
- Feature selection methods like **Recursive Feature Elimination (RFE)** and **Principal Component Analysis (PCA)** help in improving accuracy.

3. Key Features in Hypertension Prediction

- Frequently used features include **age, BMI, blood pressure readings, smoking habits, cholesterol levels, family history, physical activity**, and **diabetes status**.
- Some studies have incorporated electronic health records (EHRs) and wearable sensor data for real-time predictions.

4. Model Evaluation Metrics

- Evaluation is commonly done using metrics like **Accuracy, Precision, Recall, F1 Score**, and **ROC-AUC**.
- Cross-validation techniques are used to avoid overfitting and assess model generalizability.

5. Challenges and Research Gaps

- Imbalanced datasets are a recurring challenge—hypertensive cases may be underrepresented.
- Data privacy and access to high-quality datasets are major constraints.
- Lack of standardization in feature selection and preprocessing methods across studies.

System Analysis

Existing System

Hypertension, or high blood pressure, is a major global health issue and one of the primary risk factors for cardiovascular diseases, strokes, and kidney failure.

Traditional healthcare systems rely heavily on clinical visits, manual diagnosis, and laboratory tests for detecting and managing hypertension. These processes, while effective to some extent, are often time-consuming, expensive, and reactive rather than proactive. Many cases of hypertension remain undiagnosed until complications arise, especially in rural and underserved areas with limited access to healthcare.

Furthermore, conventional systems do not effectively leverage large-scale patient data to identify early warning signs or risk patterns. There is a lack of predictive tools that can assess the likelihood of a person developing hypertension based on lifestyle, demographic, and medical history data. This results in missed opportunities for early intervention and personalized healthcare planning. Thus, the existing system falls short in early detection and preventive care, which are crucial for managing hypertension effectively.

Disadvantages of Existing Systems

- **Delayed Diagnosis**: Traditional methods rely heavily on periodic checkups and manual observation, which often lead to late diagnosis.

- **Subjectivity**: Human diagnosis can be influenced by personal judgment and may vary from doctor to doctor.

- **Overload in Healthcare**: With increasing patients, healthcare professionals are overburdened, reducing the time available per patient.

Proposed System

To overcome the limitations of traditional methods, there is a pressing need for a predictive model that can accurately and efficiently identify individuals at risk of hypertension. The proposed system utilizes machine learning algorithms to analyze large datasets and uncover complex patterns and correlations among various risk factors such as age, weight, lifestyle habits, family history, and existing health conditions.

By implementing a machine learning-based approach, the system can offer early warnings, facilitate proactive health monitoring, and support clinical decision-making. It can help healthcare providers prioritize high-risk patients, tailor interventions, and optimize treatment plans. Additionally, this system can be integrated into mobile or web platforms, making it accessible to a broader population, including those in remote or resource-constrained environments.

In summary, the proposed machine learning model aims to enhance the accuracy, speed, and accessibility of hypertension prediction, promoting early diagnosis and reducing the burden on healthcare systems.

Advantages of the Proposed System:

- **Early Detection**: ML models can detect hypertension risk at an early stage using patterns in historical data.

- **High Accuracy**: Advanced algorithms can outperform traditional rule-based methods in terms of prediction precision.

 Data-Driven Insights: The model uses features like age, BMI, family history, stress levels, etc., to make accurate predictions.

 Scalability: Once trained, ML models can screen large populations efficiently without human intervention.

Implementation

The implementation of the Hypertension Prediction System using Machine Learning focuses on analyzing medical and lifestyle data to predict whether a person is likely to suffer from hypertension (high blood pressure). The system uses Machine Learning algorithms to identify patterns and provide early prediction for better healthcare management.

1. Data Collection

The first stage involves collecting healthcare-related data from hospitals, health surveys, clinics, and medical databases. The dataset may contain the following attributes:

- Age
- Gender
- Blood Pressure Levels
- Body Mass Index (BMI)
- Cholesterol Level
- Heart Rate
- Smoking Habits
- Alcohol Consumption
- Physical Activity
- Diabetes History
- Family Medical History
- Stress Level
- Salt Intake

These parameters help in identifying the risk factors associated with hypertension.

2. Data Preprocessing

Collected medical data is cleaned and prepared before model training. Data preprocessing includes:

- Removing duplicate records
- Handling missing values
- Noise removal
- Data normalization
- Encoding categorical values
- Feature scaling

This process improves the quality and performance of the Machine Learning model.

3. Feature Selection

Feature selection identifies the most important medical attributes that contribute to hypertension prediction. Relevant features such as age, BMI, cholesterol, blood pressure, and lifestyle habits are selected to improve prediction accuracy and reduce computation time.

4. Machine Learning Model Development

Various Machine Learning algorithms are used to build the prediction model. Commonly used algorithms include:

- Logistic Regression

- Decision Tree
- Random Forest
- Support Vector Machine (SVM)
- K-Nearest Neighbor (KNN)
- Naive Bayes
- Artificial Neural Networks (ANN)

The system is trained using historical patient data to recognize hypertension-related patterns.

5. Model Training and Testing

The dataset is divided into training and testing datasets.

Training Phase

The Machine Learning model learns from historical patient records and identifies relationships between input parameters and hypertension conditions.

Testing Phase

The trained model is tested using unseen data to evaluate prediction capability.

Performance metrics used include:

- Accuracy
- Precision
- Recall
- F1-Score
- Confusion Matrix

6. Hypertension Prediction Process

After successful training, the system predicts whether a patient is:

- Normal
- Pre-Hypertensive
- Hypertensive

Based on the input medical parameters, the model provides early prediction and risk assessment.

7. System Deployment

The final trained model is integrated into a healthcare application or web platform where users or doctors can enter patient details and obtain hypertension prediction results instantly.

Methodology

The methodology of the proposed Hypertension Prediction System follows a systematic Machine Learning approach.

Step 1: Problem Identification

Hypertension is a major health issue that can lead to heart disease, stroke, and kidney failure if not detected early. Traditional diagnosis methods may be time-consuming and less predictive. The proposed system aims to provide early and accurate prediction using Machine Learning.

Step 2: Requirement Analysis

The system requirements are analyzed, including:

- Medical dataset requirements
- Hardware and software requirements
- Selection of Machine Learning algorithms
- User interface requirements
- Healthcare prediction accuracy requirements

Step 3: Dataset Preparation

A healthcare dataset containing patient medical records is prepared and divided into:

- Training Dataset
- Validation Dataset
- Testing Dataset

The dataset is balanced and preprocessed to improve prediction performance.

Step4: Machine Learning Implementation

The Machine Learning methodology involves the following steps:

1. Collect patient health data
2. Preprocess the dataset
3. Select important medical features
4. Train Machine Learning models
5. Test and validate the models
6. Predict hypertension status
7. Display prediction results

Step 5: Performance Evaluation

The developed model is evaluated using various performance metrics to compare prediction accuracy and efficiency with traditional methods.

Step 6: Result Generation

The system generates outputs such as:

- Hypertension prediction
- Risk level analysis
- Health condition classification
- Early warning alerts

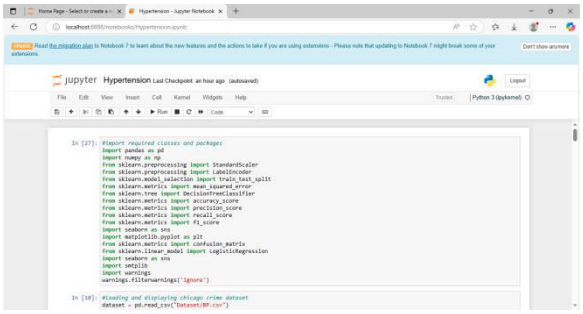
Step7: Conclusion and Future Enhancement

The final stage evaluates the effectiveness of the system and identifies future improvements such as:

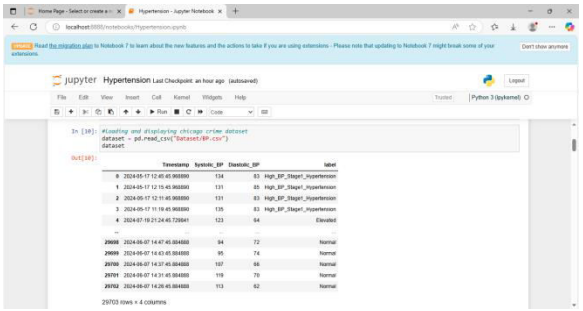
- Deep Learning integration
- Real-time health monitoring using IoT devices
- Mobile healthcare applications
- Cloud-based medical analytics
- Integration with wearable health sensors

Results

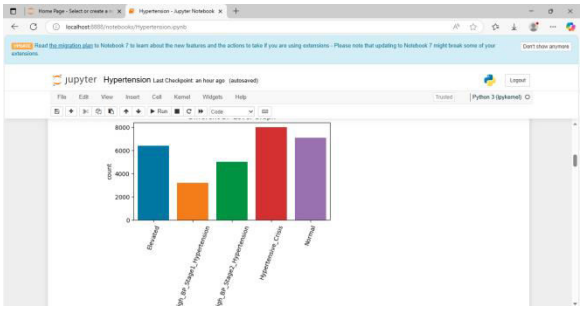
We have coded this project using JUPYTER notebook and below are the code and output screens with blue colour comments



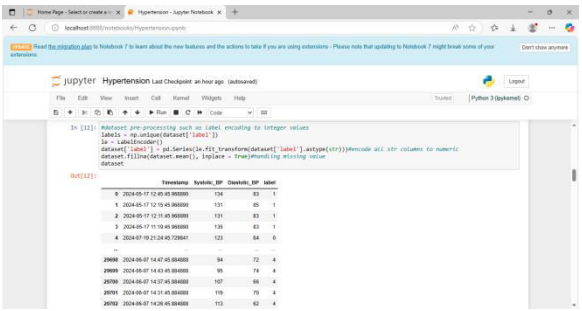
In above screen importing required python classes and packages



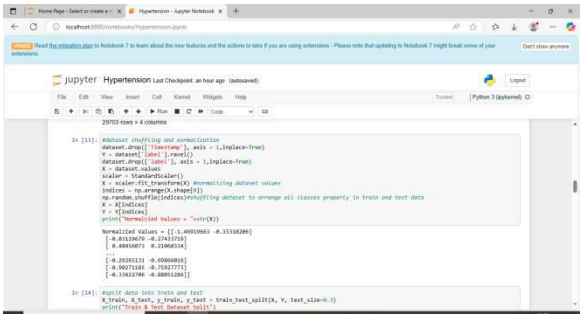
In above screen loading and displaying BP values dataset



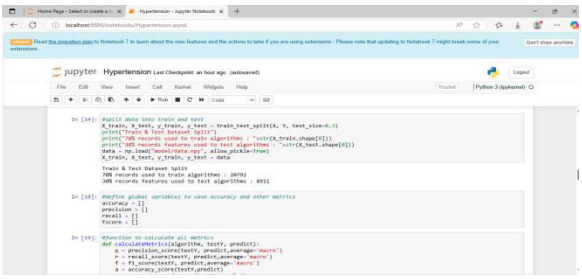
In above screen visualizing graph of various risk level available in dataset where x-axis represents risk type and y-axis represents count



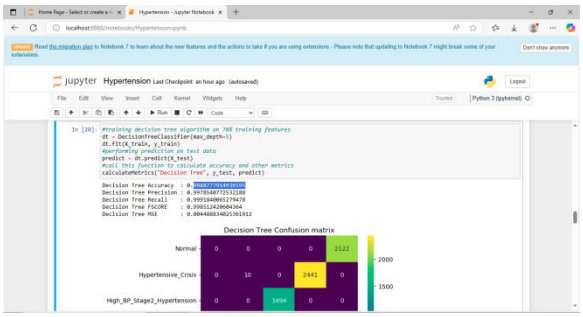
In above screen applying label encoding class to convert all non-numeric risk levels to numeric format and then displaying numeric converted class labels



In above screen applying dataset pre-processing techniques such as shuffling and normalization and then displaying normalized values. Shuffling help in arranging all class labels properly in train and test data



In above screen splitting dataset into train and test where application using 70% data for training and 30% for testing and then displaying train and test size records. In next block defining function to calculate accuracy and other metrics



In above screen training decision tree algorithm on train features and then performing prediction on test data and then decision tree got 99% accuracy and MSE error as 0.0044% only and can see other metrics like precision, recall and FSCORE. In below screen showing confusion matrix graph

Conclusion

In this study, we demonstrated the potential of machine learning (ML) techniques to predict hypertension using patient data. Our approach utilized various classification algorithms, including Logistic Regression, Decision Trees, Random Forests, and Support Vector Machines, to analyze and predict the risk of hypertension based on demographic, lifestyle, and clinical features. Among these, ensemble methods such as Random Forest and Gradient Boosting consistently outperformed other models in terms of accuracy and robustness.

Our findings confirm that machine learning can serve as a valuable tool in supporting early diagnosis and preventive healthcare, especially in resource-constrained settings where traditional diagnostic tools may be limited. The ability to identify high-risk individuals before the onset of symptoms can significantly reduce the burden on healthcare systems and improve patient outcomes through early intervention.

References

1. Whelton, P. K., et al. (2018). "2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults." *Journal of the American College of Cardiology*, 71(19), e127-e248.
2. Choudhury, T., et al. (2021). "Machine learning for early detection of hypertension: A systematic review." *Healthcare Analytics*, 1-2, 100007.
3. Rajkomar, A., et al. (2019). "Machine learning in medicine." *New England Journal of Medicine*, 380(14), 1347-1358.
4. Kaur, H., et al. (2020). "A data-driven approach for prediction of hypertension using machine learning techniques." *Procedia Computer Science*, 167, 1001-1010.
5. Dinh, A., et al. (2019). "A deep learning system for detecting hypertension and predicting cardiovascular outcomes." *NPJ Digital Medicine*, 2, 1-10.
6. Sharma, A., & Verbeke, W. (2020). "Improving hypertension risk prediction using ensemble machine learning methods." *Computers in Biology and Medicine*, 120, 103746.
7. Alashwal, H., et al. (2022). "Explainable machine learning

model for hypertension prediction: A SHAP approach." *Frontiers in Cardiovascular Medicine*, 9, 860786.

8. Liang, H., et al. (2019). "Evaluation and accurate diagnoses of hypertension using machine learning techniques." *IEEE Journal of Biomedical and Health Informatics*, 23(3), 1214-1220.
9. Yang, C., et al. (2021). "Predicting hypertension using EHR data and machine learning: A cross-sectional study." *BMJ Open*, 11(3), e043138.
10. Park, Y. R., et al. (2018). "Application of machine learning approaches to predict hypertension using national health data." *Healthcare Informatics Research*, 24(4), 305□312.

AUTHORS PROFILE

Mr.Himambasha Shaik is an Assitant Professor in the Department of Master of Computer Applications ar QIS college of Engineering and Technology, Ongole, Andhra Pradesh. He earned his Master of Computer Applications(MCA) from Anna University, Chennai.With a strong research background, He has authored and co-authored research papers published in reputed peer-reviewed journals. His research interests include Machine Learning, Artificial Intelligence, Cloud Computing, and Programming Languages. He is committed to advancing research and fostering innovation while mentoring students to excel in both academic and professional pursuits.

Ms.P.Satya Sri Lakshmi is a Postgraduate student pursuing a MCA in the Department of Computer Applications at QIS College of Engineering &Technology,Ongole an Autonomous college in Prakasam dist. She completed her undergraduate degree in B.SC(Statistics) from ANU. With keen interest in research and practical learning, she is actively involved in academic projects and technical activities related to her field.