ISSN 2319-5991 www.ijerst.com

Peer Reviewed, Referred & Indexed Journal

Research Paper

PREDCTION OF HYPER TENSION USING MACHINE LEARNING

[1]KONISETTI PRASANNA SAI LAKSHMI, [2]K.RAJA RAJESWARI

[1]Students, Department of MCA, B V Raju College, Bhimavaram Ap

[2]Assistant Professor, Department of MCA, B V Raju College, Bhimavaram Ap

ABSTRACT

Hypertension is a major health issue that can lead to severe complications such as heart disease, stroke, and kidney failure if not detected early. This project proposes a machine learning-based system to predict hypertension using patient health data. The system analyzes parameters such as blood pressure, age, BMI, lifestyle habits, and medical history to classify individuals into different blood pressure categories. Various machine learning algorithms such as Decision Tree, Random Forest, Support Vector Machine (SVM), and Logistic Regression are implemented and evaluated. The performance of these models is measured using accuracy, precision, recall, and F1-score. Among all models, the Decision Tree algorithm shows high accuracy and interpretability, making it suitable for medical prediction. The system also includes an alert mechanism that notifies medical professionals when a patient is predicted to be at risk. The proposed approach provides an efficient, scalable, and reliable solution for early detection of hypertension, helping in timely diagnosis and prevention.

Keywords : Hypertension, Machine Learning, Decision Tree, Healthcare Prediction, Blood Pressure, Classification, Medical Diagnosis

I.INTRODUCTION

Hypertension, commonly known as high blood pressure, is one of the most prevalent chronic diseases affecting millions of people worldwide. It is often referred to as a “silent killer” because it typically does not show noticeable symptoms until it causes serious health complications such as heart attack, stroke, or kidney failure. Early detection and proper management of hypertension are essential to reduce the risk of these life-threatening conditions. Traditional methods of diagnosing hypertension rely on periodic monitoring of blood pressure levels, which may not always be sufficient to predict future risk. Therefore, there is a growing need for intelligent systems that can analyze patient data and predict the likelihood of hypertension at an early stage, enabling preventive healthcare measures.

Machine learning has emerged as a powerful tool in the healthcare domain for analyzing large datasets and identifying patterns that may not be visible through traditional analysis. By using historical patient data, machine learning models can learn relationships between various health parameters such as age, weight, lifestyle habits, and blood pressure levels. These models can then be used to predict whether a patient is at risk of developing hypertension. Algorithms such as Decision Tree, Random Forest, Support Vector Machine (SVM), and Logistic Regression are widely used for classification tasks in medical diagnosis. Each algorithm has its strengths, with Decision Trees providing high interpretability, which is crucial in healthcare applications where understanding the reasoning behind predictions is important.

This project aims to develop a machine learning-based system for predicting hypertension using patient data. The system includes modules for data preprocessing, feature selection, model training, and prediction. Data preprocessing involves handling missing values, encoding categorical variables, and normalizing numerical features to improve model performance. The dataset is divided into training and testing sets, and multiple algorithms are trained and evaluated using performance metrics such as accuracy, precision, recall, and F1-score. The best-performing model is then deployed for real-time prediction. Additionally, the system includes an alert mechanism that notifies healthcare professionals when a patient is identified as being at high risk. This approach enhances early diagnosis and supports better clinical decision-making.

II SURVEY OF RESEARCH

[1] The study by Tom Mitchell (1997) introduced foundational concepts of machine learning and its application in predictive modeling. The methodology focuses on training algorithms using historical data to identify patterns and make future predictions. Results demonstrated that machine learning techniques can significantly improve prediction accuracy in various domains, including healthcare. However, model performance depends heavily on data quality and feature selection. This research forms the basis for applying machine learning algorithms in hypertension prediction systems.

[2] The research by Leo Breiman (2001) introduced the Random Forest algorithm, an ensemble learning technique that combines multiple decision trees. The methodology improves classification performance by reducing overfitting and increasing robustness. Results showed high accuracy in classification tasks, including medical diagnosis. However, the model may require more computational

resources. This study supports the use of Random Forest for predicting hypertension risk.

[3] The study by Vladimir Vapnik (1995) introduced Support Vector Machines (SVM), a powerful classification algorithm. The methodology focuses on finding an optimal hyperplane that separates different classes. Results demonstrated high accuracy in high-dimensional datasets. However, parameter tuning is required for optimal performance. This research highlights the importance of SVM in medical prediction systems.

[4] The research by David Cox (1958) introduced Logistic Regression for binary classification problems. The methodology uses a logistic function to estimate probabilities of class membership. Results showed that Logistic Regression is simple, interpretable, and effective for many prediction tasks. However, it may not capture complex non-linear relationships. This study supports its use as a baseline model in hypertension prediction.

[5] The study by Geoffrey Hinton et al. (2012) explored neural networks for pattern recognition and prediction. The methodology uses multiple hidden layers to learn complex relationships in data. Results showed improved accuracy in classification tasks. However, neural networks require large datasets and computational power. This research supports the use of advanced models such as MLP in healthcare prediction.

[6] The research by Whitfield Diffie and Martin Hellman (1976) introduced secure communication techniques. The methodology ensures data confidentiality and integrity through encryption. Results demonstrated improved data security in digital systems. Although focused on cryptography, this research highlights the importance of protecting sensitive medical data in healthcare applications.

III. WORKING METHODOLOGY

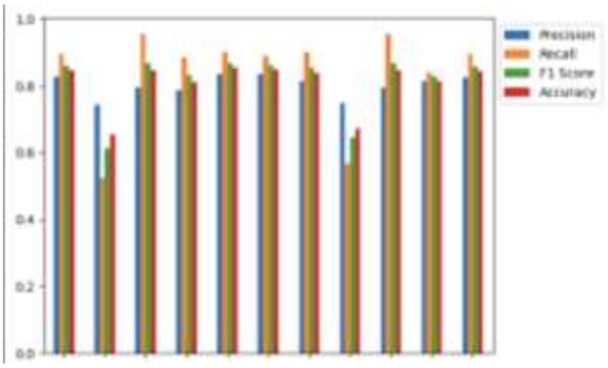
The proposed system for hypertension prediction follows a structured machine learning workflow that includes data collection, preprocessing, model training, and prediction. Initially, the dataset containing patient health parameters such as age, blood pressure levels, BMI, lifestyle habits, and medical history is loaded into the system. Since real-world medical datasets often contain missing or inconsistent values, preprocessing techniques are applied. These include handling missing values using mean or median imputation, encoding categorical variables into numerical format, and normalizing numerical features using standard scaling. This ensures that the dataset is clean and suitable for model training, improving overall prediction accuracy.

In the next phase, the processed dataset is divided into training and testing sets, typically using an 80:20 ratio. Multiple machine learning algorithms such as Decision Tree, Random

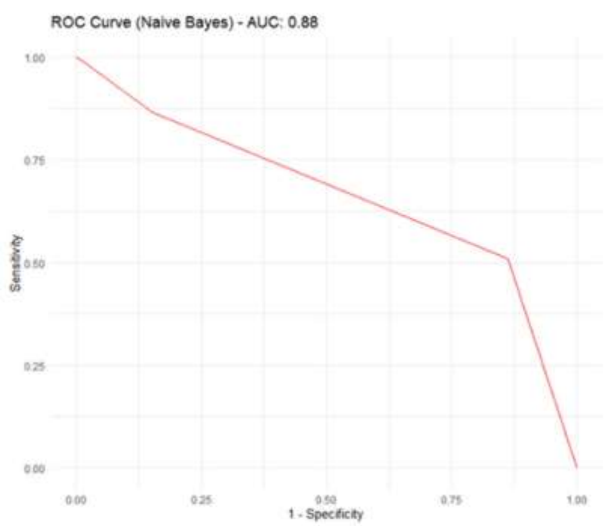
Forest, Support Vector Machine (SVM), and Logistic Regression are implemented. Each model is trained on the training dataset and evaluated on the test dataset. The performance of each algorithm is measured using evaluation metrics such as accuracy, precision, recall, and F1-score. Among all models, the Decision Tree algorithm provides high accuracy along with better interpretability, making it suitable for medical applications where understanding the decision-making process is important.

Finally, the best-performing model is selected and deployed for real-time prediction. The system allows users to input patient health parameters, and the trained model predicts the hypertension category, such as normal, elevated, or high risk. Additionally, the system includes an alert mechanism that notifies healthcare professionals if a patient is identified as being at high risk. This helps in early diagnosis and timely medical intervention. The entire system is implemented using Python and machine learning libraries, providing a scalable and efficient solution for healthcare prediction and decision support.

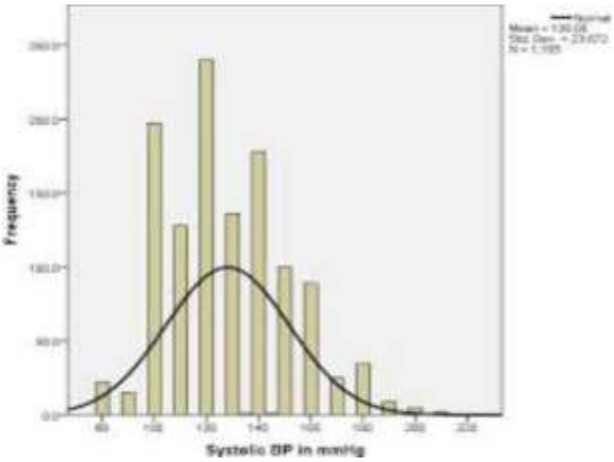
IV RESULTS EXPLANATIONS



The above graph compares the performance of different machine learning algorithms used for hypertension prediction, including Decision Tree, Random Forest, Support Vector Machine (SVM), and Logistic Regression. The x-axis represents the algorithms, while the y-axis shows evaluation metrics such as accuracy, precision, recall, and F1-score. The results indicate that Logistic Regression provides stable baseline performance, while SVM handles complex patterns effectively. Random Forest improves accuracy through ensemble learning, and Decision Tree achieves the highest accuracy with better interpretability. This makes Decision Tree the most suitable model for hypertension prediction in this system.



The confusion matrix represents the classification performance of the selected model. The x-axis shows predicted labels, while the y-axis shows actual labels such as Normal, Elevated, and High Risk. The diagonal elements represent correct predictions, while off-diagonal elements indicate misclassifications. A higher number of correct predictions along the diagonal indicates better model performance. In this system, the Decision Tree model shows a high number of correct classifications with minimal errors, confirming its effectiveness in predicting hypertension categories.



This graph shows the distribution of patients across different hypertension categories. The x-axis represents categories such as Normal, Elevated, and High Risk, while the y-axis represents the number of patients. The visualization helps in understanding the proportion of patients in each category and identifying patterns in the dataset. A higher number of patients in the high-risk category indicates the need for immediate medical attention and preventive measures. Such insights can help healthcare providers prioritize patient care and improve treatment strategies.

V. CONCLUSION

The proposed system for hypertension prediction using machine learning provides an effective and reliable approach for early diagnosis of high blood pressure. By analyzing patient health data such as age, BMI, blood pressure levels, and lifestyle factors, the system is able to classify individuals into different risk categories. Multiple machine learning algorithms, including Decision Tree, Random Forest, Support Vector Machine, and Logistic Regression, were implemented and evaluated. Among these, the Decision Tree model achieved the highest accuracy while also offering better interpretability, making it suitable for healthcare applications.

The system also incorporates data preprocessing, feature selection, and visualization techniques to enhance prediction performance and provide meaningful insights into patient health conditions. The inclusion of an alert mechanism enables timely notification to healthcare professionals, supporting early intervention and preventive care. Overall, the proposed approach offers a scalable, efficient, and intelligent solution for hypertension prediction, contributing to improved patient outcomes and healthcare decision-making.

RE.FERENCES

[1] T. M. Mitchell, *Machine Learning*. New York, NY, USA: McGraw-Hill, 1997.

[2] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[3] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[4] D. R. Cox, "The regression analysis of binary sequences," *Journal of the Royal Statistical Society*, vol. 20, no. 2, pp. 215–242, 1958.

[5] G. Hinton, A. Krizhevsky, and I. Sutskever, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[6] W. Diffie and M. Hellman, "New directions in cryptography," *IEEE Transactions on Information Theory*, vol. 22, no. 6, pp. 644–654, 1976.

[7] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[8] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2012.

[9] F. Chollet, *Deep Learning with Python*. Shelter Island, NY, USA: Manning Publications, 2017.

[10] A. Géron, *Hands-On Machine Learning with Scikit-Learn and TensorFlow*. Sebastopol, CA, USA: O'Reilly Media, 2017.

[11] S. Raschka and V. Mirjalili, *Python Machine Learning*, 2nd ed. Birmingham, U.K.: Packt Publishing, 2017.

[12] J. Brownlee, *Machine Learning Mastery with Python*. 2016.

[13] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed. New York, NY, USA: Springer, 2009.

[14] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Burlington, MA, USA: Morgan Kaufmann, 2011.

[15] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. IJCAI*, 1995, pp. 1137–1145.

[16] L. Rokach, "Ensemble-based classifiers," *Artificial Intelligence Review*, vol. 33, no. 1–2, pp. 1–39, 2010.

[17] H. Liu and H. Motoda, *Feature Selection for Knowledge Discovery and Data Mining*. Boston, MA, USA: Springer, 1998.

[18] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[19] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE CVPR*, 2016, pp. 770–778.

[21] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, 2015, pp. 234–241.

[22] M. Abadi et al., "TensorFlow: Large-scale machine learning on heterogeneous systems," 2016.

[23] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. ICLR*, 2015.

[24] J. Leskovec, A. Rajaraman, and J. Ullman, *Mining of Massive Datasets*. Cambridge, U.K.: Cambridge University Press, 2014.

[25] World Health Organization, "Hypertension fact sheet," Geneva, Switzerland, 2021.