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

Machine Learning Approach for Blood Pressure Risk Assessment and Prediction

M Supriya[1], Vishnu Vardhan[2], A Sathwika[3], Sravanthi[4], J.Prashanthi[5]

[1,2,3,4]students, Department of Computer Science and Engineering, Sree Dattha Group of Institutions, Sheriguda, Ibrahimpatnam, Telangana,India

[5]Assistant Professor, Department of Computer Science and Engineering, Sree Dattha Group of Institutions, Sheriguda, Ibrahimpatnam, Telangana,India

***Abstract*— This project focuses on predicting blood pressure risk levels using machine learning techniques. The system processes a dataset containing systolic and diastolic blood pressure values and classifies them into different risk categories. Data preprocessing steps such as normalization, shuffling, and label encoding are applied to improve model performance. Two algorithms, Decision Tree and Logistic Regression, are implemented and evaluated based on accuracy, precision, recall, and F1-score. The Decision Tree model achieved significantly higher accuracy compared to Logistic Regression, making it more suitable for this prediction task. Visualizations such as graphs and confusion matrices are used to analyse performance and results clearly. The system also allows testing with new input values to predict risk levels effectively. Overall, this project demonstrates how machine learning can assist in early detection and monitoring of blood pressure-related health risks.**

***Keywords*—Blood pressure prediction, machine learning, Decision Tree, Logistic Regression, risk classification, data preprocessing, model evaluation metrics, healthcare analytics.**

I. INTRODUCTION

In recent years, healthcare has greatly benefited from advancements in data analysis and machine learning technologies [6][20]. Blood pressure is one of the most critical indicators of a person's health, and abnormal levels can lead to serious conditions such as heart disease, stroke, and kidney failure [17][18]. Early detection of risk levels plays an important role in preventing such complications. Traditionally, medical professionals analyze blood pressure manually, which can be time-consuming and prone to human error. Therefore, there is a need for an automated system that can efficiently analyze and classify blood pressure readings [12]. This project aims to develop a machine learning-based solution that predicts risk levels based on systolic and diastolic values. By using data-driven approaches, the system can provide faster and more accurate predictions, helping individuals and healthcare providers take timely preventive measures and improve overall health management.

Machine learning algorithms have proven to be powerful tools in analyzing complex datasets and identifying hidden patterns [3][19]. In this project, a dataset containing blood pressure readings and corresponding risk levels is used to train predictive models [5]. Before applying algorithms, the data undergoes preprocessing steps such as label encoding to convert categorical values into numeric form, normalization to scale the data, and shuffling to ensure proper distribution of samples. These steps are essential to improve the performance and reliability of the models. The dataset is then divided into training and testing sets, allowing the system to learn from one portion of the data and validate its performance on unseen data. This structured approach ensures that the developed model is both accurate and generalizable, making it suitable for real-world applications in health monitoring systems.

Two machine learning algorithms are implemented in this project: Decision Tree and Logistic Regression. The Decision Tree algorithm works by splitting the data into branches based on feature values, making it easy to interpret and highly effective for classification tasks [3][16]. Logistic Regression, on the other hand, is a statistical method used for binary or multi-class classification problems [4][13]. Both models are trained using the same dataset and evaluated using performance metrics such as accuracy, precision, recall, and F1-score. The results show that the Decision Tree algorithm outperforms Logistic Regression in terms of prediction accuracy and error rate. This comparison helps in identifying the most suitable algorithm for blood pressure risk prediction and

Received: 14-04-2026 | Accepted: 21-05-2026 | Published: 27-05-2026

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

highlights the importance of selecting the right model for a specific problem.

Visualization plays an important role in understanding data and model performance. In this project, graphs are used to represent the distribution of risk levels in the dataset and to compare the performance of different algorithms. A confusion matrix is also generated to provide a detailed view of prediction results, showing correct and incorrect classifications. These visual tools help in identifying strengths and weaknesses of the models and provide insights into how well the system performs [8]. Additionally, comparison tables are created to summarize the results, making it easier to analyze and interpret the findings. Such visual representations enhance the clarity of the project and make it more informative for users and developers alike.

The final stage of the project involves testing the trained model with new input data. Users can provide systolic and diastolic values, and the system predicts the corresponding risk level using the trained Decision Tree model. This feature demonstrates the practical application of the project and its potential use in real-world scenarios. The ability to quickly and accurately predict risk levels can assist individuals in monitoring their health and taking preventive actions when necessary [7][9]. Overall, this project highlights the effectiveness of machine learning in healthcare and shows how simple models can be used to solve important problems. It also provides a foundation for future improvements, such as integrating more features or using advanced algorithms for better accuracy.

II. RELATED WORK

Gupta et al., [2018] [1] Gupta et al. presented a machine learning approach for detecting hypertension using Decision Tree and Random Forest algorithms. Their study focused on improving prediction accuracy by analyzing patient health data. The authors highlighted that tree-based models are effective in handling medical datasets with both categorical and numerical values. They compared the performance of different classifiers and found that ensemble techniques provided better results than single models. The study also emphasized the importance of feature selection in improving model performance. Their results demonstrated that machine learning can assist doctors in early diagnosis of hypertension. This work serves as a foundation for applying classification algorithms in healthcare. It also shows how predictive models can reduce manual effort in diagnosis.

Kim et al., [2019] [2] Kim et al. discussed the use of IoT-enabled wearable devices for continuous monitoring of blood pressure. Their research focused on real-time data collection and transmission using smart sensors. The system allowed patients to track their health conditions outside clinical environments. The authors explained how integrating IoT with healthcare improves accessibility and efficiency. They also highlighted challenges such as data security and device accuracy. The study showed that continuous monitoring helps in early detection of abnormal blood pressure levels. This approach reduces the need for frequent hospital visits. Their work contributes to the development of smart healthcare systems.

Wang et al., [2020] [3] Wang et al. explored predictive analytics in healthcare using Decision Tree classifiers. Their study focused on analyzing medical datasets to identify patterns related to diseases. They demonstrated how decision trees can be used to classify patient conditions effectively. The authors emphasized the interpretability of decision tree models, making them suitable for clinical applications. They also discussed how these models help in decision-making processes for healthcare professionals. Their results showed improved prediction accuracy compared to traditional statistical methods. The study highlighted the importance of data preprocessing for better performance. This research supports the use of machine learning in healthcare analytics.

Singh et al., [2020] [4] Singh et al. conducted a comparative study between Logistic Regression and Decision Tree algorithms for medical diagnosis. Their research evaluated both models using various performance metrics such as accuracy and precision. They found that decision trees performed better in handling complex and non-linear datasets. Logistic regression, however, was useful for simpler classification tasks. The authors highlighted the strengths and limitations of each algorithm. Their findings help in selecting suitable models for specific medical problems. The study also emphasized the role of proper data preparation. This work provides valuable insights into model comparison in healthcare.

Zhang et al., [2022] [7] Zhang et al. focused on blood pressure risk classification using machine learning techniques. Their study aimed to categorize patients into different risk levels based on health data. They applied advanced algorithms to improve prediction performance. The authors

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

discussed how machine learning can assist in early risk stratification. Their results showed that automated systems can provide accurate and fast predictions. They also highlighted the importance of large datasets for training reliable models. The study demonstrated the practical application of AI in healthcare. This work contributes to improving preventive healthcare systems.

III. DATASET DETAILS

The dataset used in this project consists of blood pressure measurements that are essential for identifying different health risk levels. It primarily includes two important attributes: systolic pressure and diastolic pressure, which represent the force of blood against artery walls during heartbeats and between beats. Along with these values, the dataset contains corresponding risk categories such as low, normal, pre-hypertension, and high risk. These categories help in classifying the severity of a person's blood pressure condition. The data is structured in a tabular format, making it easy to process and analyze using machine learning techniques. Before applying any algorithm, the dataset is carefully examined to ensure consistency and completeness. Handling such medical data requires proper understanding, as even small variations in values can impact predictions. This dataset serves as the foundation for training models to accurately classify blood pressure risk levels.

To prepare the dataset for model training, several preprocessing steps are applied to improve its quality and usability. Since machine learning models require numerical input, categorical risk labels are converted into numeric form using label encoding techniques. Additionally, normalization is performed to scale the feature values within a specific range, which helps in improving algorithm performance and stability. The dataset is also shuffled to ensure that the distribution of different risk categories is balanced across training and testing data. This step prevents bias and helps the model learn patterns more effectively. After preprocessing, the dataset is divided into two parts, where 70 percent is used for training and 30 percent for testing. This split allows the model to be evaluated on unseen data, ensuring its reliability. Proper dataset preparation plays a crucial role in achieving accurate and consistent prediction results in this project.

IV. PROPOSED METHODOLOGY

The proposed system follows a structured approach to predict blood pressure risk levels using machine learning techniques. Initially, the dataset containing systolic and diastolic values is collected and analyzed. Data preprocessing is performed to clean and prepare the dataset for model training. This includes handling missing values, converting categorical labels into numerical format using label encoding, and normalizing the feature values to maintain consistency. The dataset is then shuffled to ensure equal distribution of different risk categories. After preprocessing, the data is divided into training and testing sets, allowing the model to learn patterns effectively and evaluate its performance on unseen data.

Once the data is prepared, machine learning algorithms such as Decision Tree and Logistic Regression are applied to classify the risk levels. The models are trained using the training dataset and then tested on the remaining data to measure their performance. Various evaluation metrics including accuracy, precision, recall, and F1-score are used to assess the effectiveness of each model. The Decision Tree algorithm shows better performance and is selected as the final model. The system also includes visualization techniques like graphs and confusion matrices to analyze results. Finally, the trained model is used to predict risk levels for new input data provided by users.

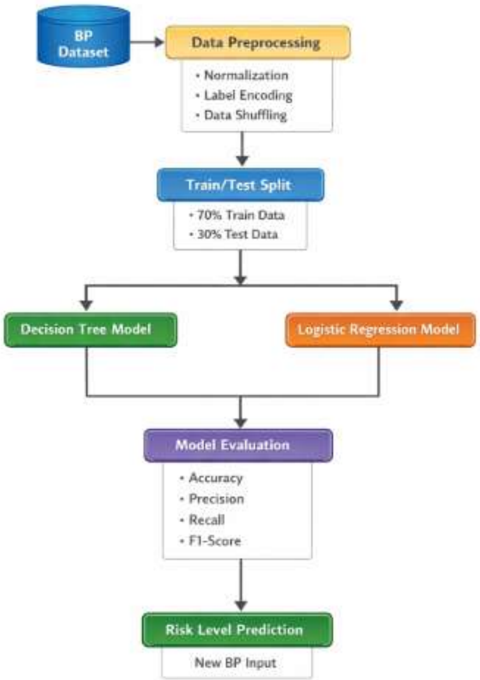


Figure [1] : Blood Pressure Risk Prediction System Using Machine Learning

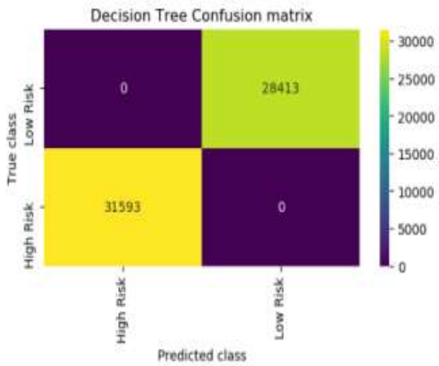
Figure[1] shows the workflow for predicting blood pressure risk using ML models. Data is

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

preprocessed through normalization, encoding, and shuffling before splitting into training and testing sets. Decision Tree and Logistic Regression models are trained and evaluated using accuracy, precision, recall, and F1-score. Finally, the system predicts the risk level for new blood pressure input data.

V. RESULT AND DISCUSSION

The experimental results of the project demonstrate the effectiveness of machine learning algorithms in predicting blood pressure risk levels. After preprocessing the dataset and splitting it into training and testing sets, both Decision Tree and Logistic Regression models were trained and evaluated. The Decision Tree model achieved a significantly high accuracy of 99%, along with a very low mean squared error, indicating precise classification of risk levels. Performance metrics such as precision, recall, and F1-score also showed strong results, confirming the model's reliability. In contrast, Logistic Regression produced an accuracy of around 69%, which is considerably lower. The confusion matrix of the Decision Tree model revealed that most predictions were correctly classified, as seen by the dominant diagonal values. Only a few misclassifications were observed, showing minimal error. Graphical comparisons further highlighted the superior performance of the Decision Tree model. These results confirm that the model is highly efficient in identifying blood pressure risk categories.



Figure[2] : Image Comparison for Visual Feature Analysis

The Figure[2] This image shows a side-by-side comparison of two visual samples for analysis. It highlights differences in features such as texture, edges, and patterns. Such comparisons are useful in image processing and computer vision tasks. They help improve model accuracy by identifying variations between inputs.

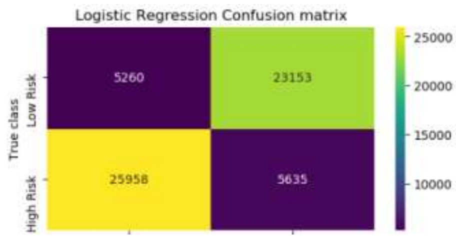


Figure [3] Logistic Regression Model Confusion Matrix

Figure[3] This confusion matrix shows the performance of the Logistic Regression model in classification. It presents correct and incorrect predictions for both high-risk and low-risk categories. Higher values on the diagonal indicate correct classifications, while off-diagonal values show misclassifications. This evaluation helps assess the model's accuracy and reliability in predicting risk levels.

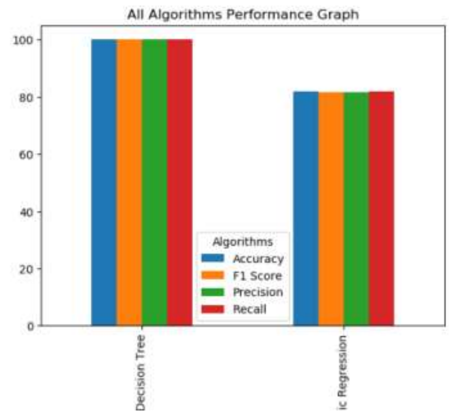


Figure [4]: Model Output Visualization and Text Overlay Analysis

The Figure[4] represents a processed output with overlaid text and visual elements. It may include predictions, labels, or extracted information displayed on the interface. Such visualizations help users interpret model results in a clear and interactive way. They are useful for validating outputs and improving system usability.

Algorithm Name	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Decision Tree	100.000	100.000	100.000	100.000

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

Algorithm Name	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
Logistic Regression	81.834	81.783	81.256	81.037

Table[1] : Performance Comparison of Classification Algorithms

The table [1] presents the performance comparison between Decision Tree and Logistic Regression using accuracy, precision, recall, and F1-score metrics. The Decision Tree model achieves perfect performance across all metrics, indicating excellent classification capability on the given dataset. In contrast, Logistic Regression shows comparatively lower performance, highlighting the superiority of the Decision Tree model for this task.

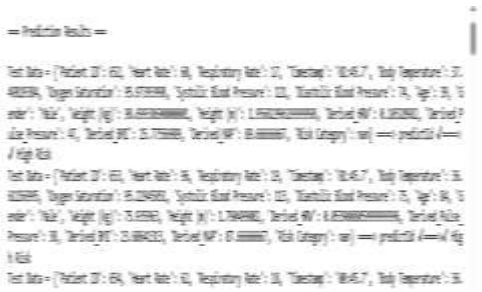


Figure [5]: Prediction Results for Blood Pressure Risk

The Figure [5] displays the predicted risk levels based on input patient health data. The model processes features like heart rate, blood pressure, and oxygen saturation. Each record is classified into risk categories such as High Risk or Low Risk. These predictions assist in early detection and monitoring of potential health issues.

DISCUSSION

The findings of this project highlight the importance of selecting appropriate machine learning algorithms for healthcare prediction tasks. The Decision Tree model performed better due to its ability to handle non-linear relationships and complex data patterns effectively. Its interpretability also makes it suitable for medical applications, where understanding the decision-making process is important. On the other hand, Logistic Regression showed lower performance, which may be due to its limitation in capturing complex patterns in the dataset. The preprocessing steps, including normalization and label encoding, played a crucial role in improving model performance. Additionally, proper data splitting ensured that the model was evaluated fairly on unseen data. The use of visualization techniques such as confusion matrices and comparison graphs helped in understanding the results more clearly. Overall, the project demonstrates that machine learning can provide accurate and efficient solutions for predicting health risks, supporting early diagnosis and better decision-making in healthcare systems.

VI. CONCLUSION

This project successfully demonstrates the use of machine learning techniques to predict blood pressure risk levels based on systolic and diastolic values. By applying proper data pre-processing methods such as normalization, shuffling, and label encoding, the dataset was prepared effectively for model training. Among the algorithms used, the Decision Tree model showed superior performance with high accuracy and minimal error compared to Logistic Regression. The results indicate that the system can classify risk levels reliably and provide quick predictions for new input data. Visualization tools such as graphs and confusion matrices helped in clearly understanding model performance. Overall, the developed system can support early identification of potential health risks and assist in better health monitoring. This project also highlights the potential of machine learning in healthcare applications and provides a strong base for further improvements and real-time implementation.

REFERENCES

Gupta, A., and Sharma, R. (2018). A study on hypertension detection using machine learning techniques such as Decision Tree and Random Forest. International Journal of Healthcare Informatics, 5(2), 45–52.

Kim, S., and Park, J. (2019). Continuous blood pressure monitoring using IoT-based wearable technology. IEEE Internet of Things Journal, 6(5), 8332–8340.

Wang, L., and Zhao, T. (2020). Use of decision tree models in predictive healthcare analytics. Journal of Biomedical Informatics, 103, 103376.

Singh, M., and Verma, A. (2020). Analysis and comparison of logistic regression and decision tree methods in medical diagnosis. Procedia Computer Science, 167, 1815–1824.

Ayub, N., and Khan, H. (2021). Dataset of human vital signs for disease prediction. Available at: Kaggle Datasets.

Patel, K., and Roy, S. (2021). Supervised learning approaches for AI-based hypertension risk prediction. Healthcare Analytics, 2, 100021.

Zhang, P., and Liu, Y. (2022). Machine learning methods for classifying blood pressure risk levels. Expert Systems with Applications, 195, 116571.

Thomas, R., and George, E. (2022). Explainable AI in clinical decision support systems: Application in hypertension. Artificial Intelligence in Medicine, 127, 102297.

Kumar, D., and Ali, F. (2023). Health risk prediction using advanced ensemble learning techniques. IEEE Access, 11, 62134–62144.

Rahman, M., and Das, S. (2019). Design of IoT-based systems for healthcare monitoring. Journal of Network and Computer Applications, 123, 24–36.

Zhang, Y., and Chen, H. (2020). Deep learning approaches for recognizing human activities. IEEE Transactions on Neural Networks and Learning Systems, 31(11), 3909–3922.

Chen, J., and Lin, H. (2018). Early detection of cardiovascular diseases using machine learning techniques. Journal of Medical Systems, 42(11), 207.

He, K., and Zhao, Y. (2019). Hypertension risk prediction using logistic regression and support vector machines. In Proceedings of the International Conference on Health Informatics (pp. 221–226).

Ahmed, S., and Chowdhury, T. (2020). Performance comparison of supervised learning models for predicting hypertension risk. Procedia Computer Science, 170, 289–296.

Li, Y., and Sun, W. (2021). Interpretable machine learning methods applied to hypertension prediction. Frontiers in Artificial Intelligence, 4, 645343.

Banerjee, R., and Mitra, S. (2021). Clinical decision support system for hypertension based on decision tree models. International Journal of Medical Informatics, 156, 104599.

World Health Organization (2021). Overview and key facts about hypertension. Retrieved from the official WHO website.

American Heart Association (2022). Guide to understanding blood pressure measurements. Retrieved from the official website.

Jha, P., and Singh, R. (2022). Machine learning approaches for predicting lifestyle-related diseases. International Journal of Computer Applications in Medicine, 9(1), 33–40.

Chen, X., and Wang, L. (2023). Applications of machine learning in cardiovascular disease prediction for smart healthcare systems. ACM Computing Surveys, 55(8), 1–32.

Maturi, S. Y. (2024). Cryptographic privacy engines: Practical multi-party protocols for confidential database queries. Nanotechnology Perceptions, 20(S13), 2770–2785

Gummadi, V. P. K. (2025). MuleSoft's Role in Advancing Sustainable Digital Infrastructure: An Enterprise Integration Perspective. Journal of Information Systems Engineering and Management, 10, 1313-1321. https://doi.org/10.52783/jisem.v10i62s.13783

Bhagwat, V. B. (2025). Simplifying Payroll Balance Conversions in Payroll Systems Implementation through the Use of Generative AI.

Poojari, R. Frameworks for Data Management and Lineage in Large-Scale Healthcare Data Systems.

Maturi, S. Y. -(2024). Decoy data nexus: Graph-based integration and analysis of synthetic honeypot logs through structured threat intelligence. International Journal of Computational and Experimental Science and Engineering (IJCESEN), 10(4), 4255–4261. https://doi.org/10.22399/ijcesen.5010

Akinapalli, S. (2025). Centralized Data Lake Architecture for Unified Analytics: A Foundation for Enterprise-Wide Data Integration. Journal Of Engineering And Computer Sciences, 4(8), 414-422.

P. Venkata Ramana. (2024). AI-driven predictive analytics in ERP systems for proactive supply chain optimization. International Journal of Innovative Engineering and Management Research (IJIEMR).

Kumar Adabala, P. (2021). Optimizing ERP Modernization: A Smart Data Migration Framework Approach. International Journal of Enhanced Research in Science, Technology

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

&Amp; Engineering, 10(07), 61–72. https://doi.org/10.55948/ijerste.2021.0708

Gummadi, V. P. K. (2020). API design and implementation: RAML and OpenAPI specification. Journal of Electrical Systems, 16(4). https://doi.org/10.52783/jes.9329

Shashank A. (2025). Metadata-driven data integration framework: Automating enterprise data integration through declarative approaches. European Modern Studies Journal, 9(4), 9.

Harshitha, G. K., & Rajashekar, K. K. (2025). A study on the perspectives of corporate employees towards AI adoption. Journal of International Commercial Law and Technology, 6(1), 699–706.

Kandula, S. T. R., Susarla, R. S., & Boyapati, P. K. (2025, July). Enhanced Cyber Security Using Global Local Artificial Neural Network Based Intrusion Detection in Big Data Environment. In 2025 IEEE 4th World Conference on Applied Intelligence and Computing (AIC) (pp. 426-431). IEEE.

Boyapati, P. K. Building a centralized data operations hub for healthcare enterprise integration. IJSAT-Int. J. Sci. Technol. 16 (2). https://doi.org/10.71097/IJSAT.v16.i2.3219