International Journal for Modern Trends in Science and Technology
Volume 12, Issue 04, pages 234-240.
ISSN: 2455-3778 online
Available online at: http://www.ijmtst.com/vol12issue04.html
DOI: https://doi.org/10.5281/zenodo.19356220


Predictive Analytics in Health Care Using Machine Learning Tools and Techniques

P.Ajaya Kumar, Chava Varshitha, Patibandla Jaini Sai, Mulakalapalli Ananya Chandrika, Shaik Afrin

Department of CSE - AI, PBR Visvodaya Institute of Technology and Science, Kavali, Andhra Pradesh, India.

To Cite this Article
P.Ajaya Kumar, Chava Varshitha, Patibandla Jaini Sai, Mulakalapalli Ananya Chandrika & Shaik Afrin (2026). Predictive Analytics in Health Care Using Machine Learning Tools and Techniques. International Journal for Modern Trends in Science and Technology, 12(04), 234-240. https://doi.org/10.5281/zenodo.19356220

Article Info
Received: 02 March 2026; Revised: 24 March 2026; Accepted: 28 March 2026.

KEYWORDS

Predictive Analytics, Machine Learning, Disease Prediction, Drug Recommendation, Clinical Decision Support System, Healthcare Informatics, Classification Algorithms.

ABSTRACT

Healthcare systems generate a large amount of patient data that can be utilized to support clinical decision making. Early identification of diseases and proper medication selection are essential to ensure effective treatment and prevent complications. This paper presents a predictive analytics–based healthcare system using machine learning techniques to assist in disease prediction and medication recommendation. The proposed system analyzes patient symptoms along with medical parameters such as age, gender, blood pressure level, cholesterol level, and sodium–potassium ratio to identify the possible disease condition and suggest suitable medication.

A supervised machine learning classification model is used to learn patterns from historical healthcare data and generate predictions for new patient inputs. After predicting the disease, the system recommends appropriate drugs using patient health attributes. The developed application provides real-time results through a user-friendly interface, helping healthcare professionals make faster and more consistent decisions. The system acts as a clinical decision support tool that can reduce diagnostic time, improve prescription accuracy, and enhance healthcare services, particularly in areas with limited medical resources.

INTRODUCTION

The expansion of digital healthcare technologies has resulted in the availability of extensive medical data that can support better patient care and treatment planning. Medical practitioners are frequently required to diagnose health conditions and determine suitable medication within a short period, and relying solely on manual evaluation and experience may lead to delayed or inconsistent decisions. To address this challenge, intelligent computational methods are needed to assist

in analyzing multiple patient factors simultaneously. Predictive analytics based on machine learning provides an effective approach by identifying patterns within healthcare data and generating meaningful predictions. In this work, a predictive healthcare system is developed that considers patient symptoms along with clinical parameters such as age, gender, blood pressure level, cholesterol level, and sodium–potassium ratio. The system predicts the probable disease condition and subsequently recommends appropriate medication, functioning as a clinical decision support tool that aims to reduce diagnosis time, improve treatment accuracy, and enhance healthcare services, particularly in resource-limited environments.

PREDICTIVE ANALYTICS

Predictive analytics is the process of examining past and present healthcare data to estimate possible future outcomes. It applies statistical analysis and data processing techniques to identify patterns within patient information and support clinical evaluation. In medical environments, predictive analytics helps practitioners interpret health records more efficiently and assists in identifying potential health conditions using available patient details.

In this work, predictive analytics is used to analyze patient symptoms and medical parameters such as age, gender, blood pressure level, cholesterol level, and sodium– potassium ratio. By evaluating these attributes, the system estimates the probable disease condition and provides guidance for treatment planning. This approach enables faster interpretation of patient data and supports timely medical attention.

MACHINE LEARNING

Machine learning is a branch of artificial intelligence that enables computer systems to learn relationships from data and generate predictions without explicit programming. A supervised learning classification model is used in this project to understand the association between patient attributes and corresponding medical outcomes. The model is trained using previously recorded healthcare data and later applied to new patient inputs.

After training, the algorithm predicts the likely disease based on reported symptoms and clinical measurements. Based on the predicted condition, the system recommends suitable medication using patient health characteristics. The application of machine learning reduces dependency on manual analysis, improves consistency in medical suggestions, and provides decision support for healthcare professionals.

OBJECTIVE

The proposed work focuses on developing a predictive healthcare support system that utilizes machine learning techniques to assist in medical analysis and treatment planning. The system examines patient symptoms and selected clinical parameters to estimate possible health conditions and provide medication suggestions. The major objectives of the study are listed below:

- To develop a predictive healthcare system using machine learning techniques.
- To analyze patient symptoms along with clinical attributes such as age, gender, blood pressure level, and cholesterol level.
- To predict possible disease conditions based on the provided patient information.
- To recommend suitable medication according to the predicted health condition.
- To support healthcare professionals in clinical decision making and reduce manual effort.
- To improve consistency and reliability in diagnosis and treatment planning.

NEED FOR STUDY

Healthcare institutions handle a large amount of patient information, and analyzing this data manually can be difficult and time-consuming for medical practitioners. Accurate identification of illnesses and appropriate treatment selection require careful consideration of multiple patient factors, and dependence only on human judgment may lead to delays or inconsistencies in clinical decisions. In many primary healthcare centers, the limited availability of experienced specialists further increases the necessity for supportive systems that can assist in medical evaluation. Predictive analytics combined with machine learning provides a practical approach to interpret patient symptoms and clinical measurements in an organized manner. A system capable of estimating probable diseases and recommending suitable medication can assist healthcare providers in making

informed decisions and improving treatment planning. Therefore, this study is required to demonstrate how data-driven techniques can enhance reliability, reduce decision complexity, and improve the accessibility and efficiency of healthcare services.

EXISTING SYSTEM

In the traditional healthcare environment, diagnosis and medication selection are primarily performed through manual evaluation by medical practitioners. Doctors analyze patient symptoms, clinical measurements, and medical history based on their knowledge and experience to determine the possible disease and prescribe treatment. Although this approach is effective in many cases, it requires significant time and depends heavily on the individual expertise of the practitioner. Variations in experience and workload may lead to inconsistencies in diagnosis and prescription decisions.

Some existing computerized systems provide electronic health record storage and basic clinical support; however, they often focus only on maintaining patient data rather than actively assisting in prediction or treatment selection. Many healthcare centers, especially in rural areas, do not have access to advanced diagnostic tools or specialist consultation, which may delay medical attention. As a result, there is a need for an intelligent system that can analyze patient information automatically and assist healthcare professionals in identifying diseases and recommending suitable medication in a more consistent and efficient manner.

Disadvantages

- Relies heavily on manual analysis and traditional statistical methods.
- Limited ability to handle large and complex healthcare datasets.
- Lower prediction accuracy due to fixed rules and assumptions.
- Time-consuming decision-making process.
- Poor performance with noisy or incomplete medical data.
- Not suitable for early disease prediction and real-time analysis.
- Lack of adaptability to changing patient conditions.

SYSTEM ARCHITECTURE

MODULES

- Data Collection Module
- Data Preprocessing Module
- Feature Selection Module
- Dataset Splitting Module
- Model Training Module
- Prediction and Classification Module
- Output and Decision Support Module

1. Data Acquisition Module:

The system begins by obtaining healthcare data from available medical datasets. The dataset contains patient information such as symptoms and clinical attributes including age, gender, blood pressure level, and cholesterol level. These records serve as the input for training and testing the predictive model.

2. Data PreProcessing Module:

The collected data is cleaned to remove incomplete or inconsistent entries. Missing values are handled and categorical attributes are converted into numerical form using encoding techniques such as label encoding. Data normalization is applied to scale attribute values so that all features contribute equally during model learning.

3. Feature Selection Module:

Important attributes influencing disease identification are selected from the dataset. Statistical analysis and correlation checking are used to retain relevant features while eliminating unnecessary attributes. This improves prediction performance and reduces computational complexity.

4. Dataset Splitting Module:
The prepared dataset is divided into training and testing sets. The training set is used for building the machine learning model, while the testing set is used to evaluate prediction accuracy on unseen patient data.

5. Model Training Module:
Supervised machine learning classification algorithms are applied to learn relationships between patient attributes and disease conditions. The Decision Tree classifier constructs a hierarchical structure based on attribute importance, while the K-Nearest Neighbor (KNN) algorithm compares new data with previously stored patient cases using distance calculation. The model is trained using the training dataset and optimized for better prediction performance.

6. Prediction and Classification Module:
When new patient data is entered, the trained model analyzes the input features and predicts the most probable disease condition. The classification output is generated based on learned patterns from historical records. After disease identification, suitable medication is recommended using the patient's clinical parameters.

7. Output and Decision Support Module:
The system displays the predicted disease along with recommended medication as the final result. The output assists healthcare professionals in understanding patient condition and supports treatment planning through data- driven recommendations.

PROPOSED SYSTEM

The proposed system is a predictive healthcare application that employs machine learning techniques to assist in identifying diseases and recommending suitable medication. The system analyzes patient symptoms by the together with clinical parameters such as age, gender, blood pressure level, and cholesterol level to estimate the probable health condition. Patient data is first collected and preprocessed to remove inconsistencies and convert attributes into a format appropriate for analysis.

A supervised classification model is trained using historical healthcare records to learn the relationship between patient characteristics and disease conditions. When new patient information is provided, the trained model predicts the likely disease and generates medication suggestions as output. This automated approach supports healthcare professionals in treatment planning, improves consistency
in clinical decisions, and provides reliable medical assistance, especially in locations where specialist consultation may not always be available.

Advantages:
- High prediction accuracy using advanced machine learning algorithms.
- Enables early disease detection to reduce complications and treatment costs.
- Supports clinical decision-making with data- driven insights.
- Effectively handles large and complex healthcare datasets.
- Adapts to new patterns in patient data for continuous learning.
- Deployable on limited-resource devices using lightweight models.
- Automates analysis, reducing manual effort and errors.
- Improved scalability through modular and flexible system architecture.

Hardware Requirements
- Processor: Intel Core i3 / i5 or equivalent processor
- RAM: Minimum 4 GB (8 GB recommended)
- Storage: At least 10 GB free disk space
- Input Device: Keyboard and mouse for entering patient details / CSV dataset
- Output Device: Monitor or display screen

Software Requirements:
- Operating System: Windows or Linux
- Programming Language: Python
- Development Environment: Jupyter Notebook or Visual Studio Code
- Machine Learning Library: Scikit-learn
- Data Processing Libraries: NumPy, Pandas
- Visualization Library: Matplotlib
- Development Environment: Jupyter Notebook or Visual Studio Code
- Machine Learning Library: Scikit-learn
- Data Processing Libraries: NumPy, Pandas

- Visualization Library: Matplotlib
- Framework (Optional): Streamlit for user interface
- Models Used: Decision Tree Classifier, K- Nearest Neighbor (KNN)
- Dataset Used: Healthcare dataset obtained from Kaggle in CSV format containing patient attributes such as age, gender, blood pressure level, cholesterol level, and related medical information.

TECHNIQUES USED IN THE PROJECT

The proposed healthcare predictive system is implemented in Python, leveraging its rich ecosystem of data science libraries. The dataset, sourced from Kaggle, is processed in CSV format. Key libraries used include:

- Pandas and NumPy for efficient data handling and preprocessing.
- Scikit-learn for building, training, and evaluating machine learning models.
- Matplotlib for visualizing data trends and model performance.

2. Data Preparation

Data cleaning is performed by removing inconsistent or incomplete records. Categorical features such as gender and medical categories are converted to numerical values using label encoding. All numerical features are normalized to ensure that each attribute contributes equally during model training.

3. Classification

The system employs supervised classification methods to predict disease conditions and recommend medication. The dataset is split into training and testing sets. Training data is used to construct predictive models, while testing data evaluates their performance.

4. Decision Tree

A Decision Tree classifier identifies the most informative attributes to split the dataset using the Gini impurity measure.

Gini impurity measure

$$Gini = 1 - \sum_{i=1}^{n}(p_i)^2$$

where *pi* is the probability of a sample belonging to class *i*, and *c*is the total number of classes. The tree structure allows classification of new patient data based on learned rules.

4.K-Nearest Neighbor (KNN)

The KNN algorithm compares new patient records with previously stored cases. Similarity is measured using Euclidean distance.

Euclidean distance:

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

The majority class among the nearest neighbors determines the predicted disease category.

5. Performance Evaluation

Model performance is evaluated using a confusion matrix. Key metrics include:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$F1\ Score = \frac{2 \times (Precision \times Recall)}{Precision + Recall}$$

These measures ensure reliable disease prediction and appropriate medication recommendations for new patients.

CONCLUSION

The Healthcare Predictive System developed in this project showcases the power of machine learning in transforming healthcare. By leveraging Decision Tree and K-Nearest Neighbor (KNN) algorithms, the system effectively analyzes patient data to predict potential disease conditions with high accuracy. Careful data preprocessing, feature encoding, and normalization ensure that the models are trained on clean, meaningful information, leading to reliable predictions.

Evaluation metrics such as accuracy, precision, recall, and F1-score demonstrate the system's consistency and dependability across diverse patient datasets. Beyond numbers, the real strength of this system lies in its ability to provide timely, actionable insights to clinicians through a user-friendly interface, supporting early disease detection and informed decision-making.

By bridging the gap between data and clinical action, this project highlights how predictive analytics can empower healthcare professionals, enhance patient outcomes, and drive proactive medical care. In essence, the system is not just a predictive tool—it is a step toward smarter, data- driven healthcare that anticipates risks before they escalate.

FUTURE ENHANCEMENT

The current Healthcare Predictive System provides accurate disease predictions and real-time insights for clinicians. However, there are several opportunities to make the system even more robust, scalable, and clinically valuable. Future enhancements can focus on expanding data sources, incorporating hybrid machine learning techniques, and improving accessibility and usability. These improvements will ensure personalized, timely, and proactive healthcare solutions.

Future Enhancements Include:

• Integration with IoT and Wearable Devices:
Collect continuous real-time patient data, such as heart rate, blood pressure, or glucose levels, to enhance prediction accuracy and monitor patient health dynamically.

• Hybrid Machine Learning Approaches:
Combine Decision Tree and KNN with other algorithms (e.g., Random Forest, Gradient Boosting) to improve prediction performance and handle more complex patient datasets.

• Support for Multimodal Data:
Incorporate structured data (lab reports, vitals) and unstructured data (doctor's notes, medical history) to generate more comprehensive and context-aware predictions.

• Personalized Risk Assessment:
Provide individualized risk scores and recommendations based on each patient's medical history, demographic data, and lifestyle factors.

• Web and Cloud-Based Deployment:
Host the system on cloud platforms to allow multi-center access, real-time updates, and secure storage of patient data.

• Automated Alerts and Notifications:
Generate real-time alerts for clinicians or patients when a high-risk condition is detected, enabling timely interventions and preventive care.

• Enhanced Visualization Tools:
Implement dashboards and graphical summaries to help clinicians easily interpret predictions and trends, supporting faster and more informed decision- making

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] Abeer Al-Nafjan, Amaal Aljuhani, Arwa Alshebel, Asma Alharbi, and Atheer Alshehri, "Artificial Intelligence in Predictive Healthcare: A Systematic Review," Journal of Clinical Medicine, vol. 14, no. 19, 2025.

[2] Mohammed Badawy, Nagy Ramadan, and Hesham Hefny, "Healthcare Predictive Analytics using Machine Learning and Deep Learning Techniques: A Survey," Journal of Electrical Systems and Information Technology, 2023.

[3] Norah Hamad Alhumaidi, Doni Dermawan, and Hanin Farhana Kamaruzaman, "The Use of Machine Learning for Analyzing Real-World Data in Disease Prediction and Management: Systematic Review," JMIR Medical Informatics, vol.13, 2025.

[4] Alok Sharma, Artem Lysenko, Shangru Jia, et al., "Advances in AI and Machine Learning for Predictive Medicine," Journal of Human Genetics, vol. 69, pp. 487– 497, 2024.

[5] Gopinath K and Anurag Shrivastava, "Predictive Analytics in Healthcare: A Machine Learning Model for Early Health Risk Detection," Frontiers in Health Informatics, vol. 13, no. 4, 2024.

[6] Nasrullah Abbasi, "Utilizing Machine Learning for Predictive Analytics: Advances in Healthcare Technology," Journal of AI-Powered Medical Innovations, vol. 1, no. 1, 2024.

[7] Kannan Vishwanatth, Savitha Satish, Chandra Sekhar Mohapatra, et al., "Predictive Analytics and Machine Learning in Disease Diagnosis: A Review of Recent Advances," Journal of Electrical Systems, Apr. 2024.

[8] Dr. Hassan Ali, "Utilizing Predictive Analytics for Lifecycle Management and Maintenance of Medical Equipment," Journal of AI in Healthcare and Medicine, May 2024.

[9] Rajendra Prasad Urukadle, "Predictive Analytics and Machine Learning in Healthcare: A Comprehensive Framework for Clinical Implementation," International Journal of Scientific Research in Computer Science, Engineering and Information Technology, vol.11, no.2, pp.702–708, Mar.2025.

[10] Md. Jawadur Rahim, Ahlina Afroz, and Omolola Akinola, "Predictive Analytics in Healthcare: Big Data, Better Decisions," International Journal of Scientific Research in Modern Technology, vol.4, no.1, 2025.

[11] R. P. Urukadle, "Predictive Analytics and Machine Learning in Clinical Healthcare Implementation," ResearchGate, 2025.

[12] N. Nurseha Amalina, K. Boateng Ofori-Amanfo, H. An, "A Multi-Head Attention Soft Random Forest for Interpretable Patient No-Show Prediction," arXiv preprint, 2025.

[13] A. A. Lamir, S. Razzagzadeh, Z. Rezaei, "A Comprehensive Machine Learning Framework for Heart Disease Prediction: Performance Evaluation and Future Perspectives," arXiv preprint, 2025.

[14] Srimanth Maddipatla, "Predictive Healthcare Analytics Using AI on Modernized Big Data Platforms," International Journal of Computational and Experimental Science and Engineering, 2026.
[15] Sara Noori Mohammad Ali,Nawzad Muhammed Ah med, "Comparing Machine Learning Models for Cardiovascular Disease," Journal of Pioneering Medical Sciences, May 2025.
[16] G. Sabeena Gnanaselvi et al., "AI and Data Analytics for Proactive Healthcare Risk Management," Journal of Neonatal Surgery, 2024.
[17] [17]Norah Hamad Alhumaidi et al., Detailed Machine Learning Disease Prediction Review, JMIR Medical Informatics, 2025.
[18] Alok Sharma et al., Predictive Models in Healthcare Analytics: Role of AI/ML, Journal of Human Genetics, 2024.
[19] Pyla Jyothi, A. Lokesh Kumar, D. Dakshayani, G. Kavya Sri, "Disease Prediction using Naive Bayes, Random Forest, Decision Tree, KNN Algorithms," i-manager Publications, 2024.