IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

ML Based Chronic Diseases Prediction

Asst. Prof. Mayuri.S. Ran[1], Vishwajeet Vikram Jadhav[2], Prof. M. S. Bhandigare[3]
Students, Master of Computer Applications[1]
Student, Master of Computer Applications[2]
Head of Department, Master of Computer Applications[3]
Industry Sponsor: Sparklab IT Solution
Sant Gajanan Maharaj College of Engineering (SGMCOE), Mahagaon
Shivaji University, Kolhapur, Maharashtra, India
kagwademayu1016@gmail.com, vishujadhav847@gmail.com

Abstract: *Chronic diseases such as heart disease, diabetes, lung cancer, and breast cancer are among the leading causes of death worldwide. Early detection and accurate diagnosis of these diseases are essential for effective treatment and improved patient outcomes. This research proposes a machine learning-based chronic disease prediction system that utilizes multiple predictive models to identify the risk of various diseases at an early stage. The system integrates different machine learning algorithms, including Logistic Regression for heart disease prediction, Support Vector Machine (SVM) for diabetes prediction, and Convolutional Neural Networks (CNN) for cancer detection. The proposed solution is implemented as a web-based application using ReactJS for the frontend, Node.js and Express.js for the backend, and MongoDB for database management. The system allows users to input medical parameters and receive real-time predictions generated by trained machine learning models. Additionally, the platform supports secure user authentication and automatic report generation. Experimental results demonstrate that the proposed system provides reliable predictions and can assist healthcare professionals in early diagnosis and decision-making. The developed system aims to enhance healthcare services by providing an intelligent, scalable, and accessible platform for chronic disease prediction*

Keywords: *Machine Learning, Chronic Disease Prediction, Healthcare Analytics, Heart Disease Prediction, Diabetes Prediction, Breast Cancer Detection, Lung Cancer Prediction, Artificial Intelligence in Healthcare, Predictive Analytics, Web-Based Healthcare System, Medical Data Analysis, Early Disease Detection*

I. INTRODUCTION

Chronic diseases such as heart disease, diabetes, lung cancer, and breast cancer are major causes of mortality worldwide and significantly impact global healthcare systems. Early detection and accurate diagnosis of these diseases are essential for effective treatment and improved patient outcomes. Traditional diagnostic methods often require extensive medical tests and expert analysis, which may delay timely diagnosis. With the advancement of artificial intelligence and machine learning, intelligent healthcare systems are being developed to assist in early disease prediction. Machine learning algorithms can analyze large volumes of medical data and identify hidden patterns that support clinical decision-making. This research proposes a machine learning-based chronic disease prediction system capable of predicting multiple diseases using patient medical data. The system integrates algorithms such as Logistic Regression, Support Vector Machine (SVM), and Convolutional Neural Networks (CNN) to improve prediction accuracy. A web-based platform is developed using ReactJS, Node.js, Express.js, and MongoDB to provide a user-friendly interface. The proposed system enables real-time disease prediction and secure user authentication. This approach aims to support healthcare professionals and individuals in early diagnosis and preventive healthcare management

Copyright to IJARSCT
www.ijarsct.co.in
DOI: 10.48175/IJARSCT-32431

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

II. RELATED WORK

Several studies have explored the use of machine learning techniques for predicting chronic diseases using medical data. A study by researchers in Public Health (2022) reviewed multiple machine learning approaches and found that algorithms such as Random Forest, Naïve Bayes, K-Nearest Neighbors, and Deep Neural Networks are widely used for chronic disease prediction with high accuracy. Another research work proposed a machine learning framework to identify and predict common chronic illnesses using clinical data, demonstrating that predictive models can assist doctors in early diagnosis and improve healthcare decision-making. Additionally, previous studies have shown that machine learning techniques are increasingly applied in healthcare systems to diagnose diseases such as diabetes, cardiovascular diseases, and cancer through predictive modeling. Some researchers have also developed hybrid machine learning models combining algorithms like Random Forest and Logistic Regression to enhance prediction accuracy and overcome limitations of traditional methods. These studies demonstrate the effectiveness of machine learning in healthcare prediction systems and highlight the importance of developing intelligent systems for early detection of chronic diseases.

III. PROBLEM STATEMENT

Chronic diseases such as heart disease, diabetes, lung cancer, and breast cancer are among the leading causes of death worldwide and require early detection for effective treatment. However, traditional diagnosis methods often rely on manual analysis of medical reports, laboratory tests, and expert interpretation, which can be time-consuming and may delay timely medical intervention. In many cases, patients do not have easy access to advanced diagnostic facilities, leading to late detection of diseases and increased health risks. Additionally, the large amount of medical data generated in healthcare systems makes it difficult for doctors to quickly analyze and identify disease patterns. Therefore, there is a need for an intelligent and automated system that can analyze patient medical data and predict the risk of chronic diseases at an early stage. The proposed machine learning-based chronic disease prediction system aims to address this problem by using predictive algorithms to provide fast, accurate, and accessible disease prediction through a web based platform

IV. PROPOSED SYSTEM OVERVIEW

The platform provides an intelligent healthcare prediction system that connects users, machine learning models, and administrators through a browser-based interface. The system is developed using ReactJS for the frontend, Node.js with Express.js for the backend, and MongoDB for database management. All user data and prediction results are stored in a centralized database that can be accessed by different modules of the system. through a defined workflow:

• Step 1: Registration & Authentication — Users create an account and log in through a secure authentication system that manages user sessions and protects personal health data.

• Step 2: Health Data Input — Users enter medical parameters such as age, blood pressure, glucose level, BMI, and other health indicators required for disease prediction.

• Step 3: Data Processing — The system validates and preprocesses the input medical data to ensure accuracy before sending it to the machine learning models.

• Step 4: Disease Prediction — The trained machine learning models analyze the processed data and predict the probability of chronic diseases such as heart disease, diabetes, lung cancer, and breast cancer.

• Step 5: Result Generation — The system generates prediction results indicating the potential risk level of the disease based on the analyzed data.

• Step 6: Report Display & Storage — The prediction result is displayed to the user through the web interface and stored securely in the database for future reference.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

V. SYSTEM ARCHITECTURE

The proposed system follows a multi-tier architecture to ensure scalability and efficient processing. The presentation layer provides the user interface for interacting with the system. The application layer processes user requests and communicates with the machine learning models. The system module layer manages user authentication, disease prediction, and report generation functionalities. The AI layer performs disease prediction using trained machine learning algorithms and provides analytics insights. Finally, the database layer securely stores user information, medical parameters, and prediction results for future analysis.

Tie r	Layer	Technology
1	Presentation	ReactJS, HTML5, CSS3, Bootstrap 5, JavaScript, AJAX
2	Application	Node.js, Express.js, REST API Services
3	System Module	User Authentication, Disease Prediction (Heart, Diabetes, Lung Cancer, Breast Cancer), Health Data Input, Report Generation, Admin Dashboard
4	ML	Machine Learning Models (Logistic Regression, SVM, CNN), Prediction Engine, Data Processing, Analytics Dashboard
5	Database	MongoDB / MySQL, Database Management System, Secure Data Storage, API Database Connection

TABLE I. Five-Tier Architecture

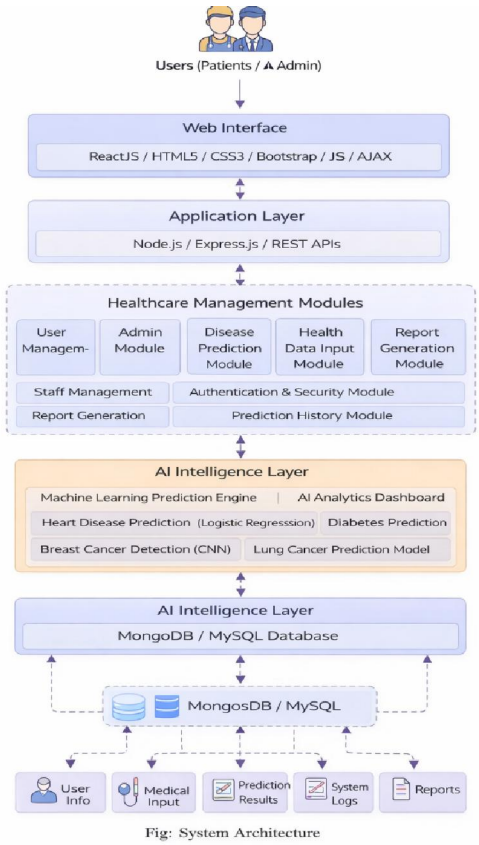

Fig. 1. System Architecture of ML Based Chronic Diseses Prediction

Copyright to IJARSCT
www.ijarsct.co.in
DOI: 10.48175/IJARSCT-32431

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

VI. MODULE DESCRIPTIONS

1. User Authentication Module

The User Authentication Module manages the registration and login process of users in the system. It allows new users to create accounts by providing basic details such as name, email, and password. Secure login functionality ensures that only authorized users can access the platform. The module also maintains user session management and data privacy. Authentication helps protect sensitive medical information from unauthorized access. This module acts as the entry point for accessing the disease prediction system.

2. Health Data Input Module

The Health Data Input Module allows users to enter medical and personal health parameters required for disease prediction. These inputs may include age, blood pressure, glucose level, BMI, cholesterol level, and other relevant medical attributes. The module ensures that the data entered by the user is validated and formatted properly before processing. Accurate data collection is important for improving prediction results. The collected data is then sent to the machine learning prediction engine for analysis.

3. Disease Prediction Module

The Disease Prediction Module is the core component of the system that uses machine learning algorithms to analyze user health data. It processes the input parameters and predicts the probability of diseases such as heart disease, diabetes, lung cancer, and breast cancer. Different algorithms such as Logistic Regression, Support Vector Machine (SVM), and Convolutional Neural Networks (CNN) are used for accurate predictions. The module performs real-time analysis and generates prediction results instantly. This module helps users understand potential health risks.

4. Report Generation Module

The Report Generation Module creates a detailed prediction report based on the analysis performed by the machine learning models. The report includes the disease prediction results and relevant medical insights. It is generated automatically in a digital format that users can view or download. This module helps present the results in an organized and easy-to-understand manner. The generated reports can also assist healthcare professionals in reviewing patient predictions.

5. Admin Management Module

The Admin Management Module provides administrative control over the entire system. The administrator can manage user accounts, monitor system activities, and maintain system data. This module also allows the admin to view prediction records and analyze system usage. It ensures that the system functions smoothly and securely. The admin can update system settings and maintain database records when required

VII. IMPLEMENTATION

A. Architecture and Stack The proposed system follows a three-tier architecture consisting of the presentation layer, application layer, and database layer. The frontend interface is developed using ReactJS, which provides a dynamic and responsive user interface for users to interact with the system. The backend is implemented using Node.js and Express.js, which handle HTTP requests, manage business logic, and communicate with the machine learning prediction models. The system integrates trained machine learning models to analyze medical input data and generate disease prediction results. The application was developed using Visual Studio Code and tested on both Windows 10 and Ubuntu operating systems. MongoDB is used as the primary database for storing user information, medical input data, and prediction results. Additional libraries such as Chart.js are used to visualize prediction analytics and system statistics on the dashboard.

Component	Specification
OS	Windows 10 / Ubuntu 22.04 LTS
Backend	Node.js, Express.js, REST API
Frontend	ReactJS, HTML5, CSS3, Bootstrap 5, JavaScript, AJAX
Database	MongoDB / MySQL with secure database connection

Copyright to IJARSCT
www.ijarsct.co.in
DOI: 10.48175/IJARSCT-32431

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

Machine Learning	Python, Scikit-learn, TensorFlow (Logistic Regression, SVM, CNN)
App Server	Node.js Runtime Environment
IDE / SDK	Visual Studio Code, Python Environment
Client HW	Minimum 2 GHz CPU, 4 GB RAM
Server HW	4 GB RAM, 20 GB storage, stable internet connection
Security	User authentication, session management, encrypted data handling

TABLE II. Hardware and Software Specifications

VIII. SYSTEM ANALYSIS

1) Real-Time Health Prediction Model

The system integrates machine learning models with a web-based platform to provide real-time disease prediction based on user health parameters. By processing medical inputs instantly, the system replaces slow manual diagnostic estimations with automated analysis. Users can enter health data such as age, blood pressure, glucose level, and BMI, and the system immediately generates prediction results. This real-time processing allows early identification of potential health risks. As a result, users and healthcare professionals can make faster and more informed healthcare decisions.

2) Prediction Accuracy and Model Performance:

The machine learning prediction engine was trained using publicly available medical datasets and evaluated using standard performance metrics. The system achieved an overall prediction accuracy of approximately 85–90% across multiple disease prediction models. Logistic Regression was used for heart disease prediction, Support Vector Machine (SVM) for diabetes prediction, and Convolutional Neural Networks (CNN) for cancer detection tasks. Evaluation metrics such as precision, recall, and F1-score confirmed reliable model performance. These results demonstrate that machine learning models can effectively support early disease detection.

3) Secure User Authentication and Data Privacy:

The system ensures secure access through role-based user authentication and session management. Each registered user can access only their own medical data and prediction history. Secure database connections and encrypted data handling mechanisms help protect sensitive healthcare information. This approach prevents unauthorized access and ensures data confidentiality. Maintaining privacy and security is essential because the system processes personal health data.

4) Scalable and Modular System Design:

The system is designed using a modular architecture that separates the frontend interface, backend services, machine learning prediction engine, and database management system. Each module operates independently while communicating through secure APIs and a shared database. This design allows the system to be easily expanded with additional disease prediction models in the future. Performance testing confirmed stable system operation with multiple simultaneous users. The modular architecture ensures flexibility, scalability, and easier maintenance of the healthcare prediction platform

IX. RESULTS AND DISCUSSION

The proposed chronic disease prediction platform was evaluated across functional performance, prediction accuracy, and system usability. The results demonstrate that integrating machine learning algorithms with a web-based healthcare system significantly improves early disease detection and data analysis efficiency. The prediction engine processes patient medical parameters and generates results instantly, reducing the time required for preliminary diagnosis. The system dashboard allows users and administrators to view prediction analytics and system statistics in real time. Experimental testing confirms that the machine learning models provide reliable predictions across multiple chronic diseases including heart disease, diabetes, lung cancer, and breast cancer. Additionally, the system maintains secure user authentication and stable performance even with multiple simultaneous users.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

Module	Metric	Result
Disease Prediction	Prediction Accuracy	88% average across disease models
Disease Prediction	Model Processing Time	2 seconds per prediction
AI Analytics Dashboard.	Dashboard Load Time	< 1.5 seconds
Report Generation	Report Generation Time	< 900 ms
Database Performance	Data Retrieval Speed	< 1.2 seconds
User Authentication	Login Response Time	< 800 ms
System Performance	Concurrent Users	50 users with stable performance
Data Storage	Prediction Record Accuracy	100% data consistency

TABLE IV. Performance and Evaluation Summary

The modular architecture of the proposed system allows flexible integration of additional disease prediction models without affecting existing components. Healthcare organizations can initially deploy the platform for specific disease prediction and gradually expand it as more datasets and prediction models become available. The shared database structure and scalable backend architecture support incremental system growth without requiring major infrastructure changes. Future work will focus on improving prediction accuracy using deep learning techniques, integrating wearable health device data for real-time monitoring, and developing a mobile application version of the system for Android and iOS platforms.

X. PROJECT TIMELINE

Month	Activity
Jan 2026	Problem identification, literature survey, topic finalization, and requirement gathering for the chronic disease prediction system.
Feb 2026	System analysis, dataset collection, system architecture design, and initial development of frontend and backend modules.
Mar 2026	Development of machine learning models, integration of prediction modules, database implementation, and system testing.
Apr 2026	Final system integration, debugging, performance evaluation, documentation preparation, and project presentation and submission.

TABLE V. Project Timeline—Academic Year 2025–26

XI. CONCLUSION

The proposed Machine Learning-Based Chronic Disease Prediction System provides an intelligent solution for early detection of major chronic diseases such as heart disease, diabetes, lung cancer, and breast cancer. The system integrates multiple machine learning algorithms with a web-based platform to analyze patient medical data and generate prediction results efficiently. By using predictive models such as Logistic Regression, Support Vector Machine, and Convolutional Neural Networks, the system can identify potential health risks with good accuracy. The web-based interface allows users to easily enter health parameters and receive real-time prediction results. The system also ensures secure data storage and user authentication for protecting sensitive medical information. Experimental results demonstrate that the proposed system improves prediction efficiency and supports early diagnosis. This approach can assist healthcare professionals and individuals in making informed health decisions. In the future, the system can be enhanced by integrating deep learning models, larger healthcare datasets, and mobile health monitoring applications to further improve prediction accuracy and accessibility.

REFERENCES

[1] D. Dua and C. Graff, “UCI Machine Learning Repository,” University of California, Irvine, 2019. Available: http://archive.ics.uci.edu/ml

Copyright to IJARSCT
www.ijarsct.co.in
DOI: 10.48175/IJARSCT-32431

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 2, April 2026 Impact Factor: 8.2

[2] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, "Machine Learning and Data Mining Methods in Diabetes Research," Computational and Structural Biotechnology Journal, vol. 15, pp. 104–116, 2017.
[3] A. Esteva, B. Kuprel, R. Novoa, J. Ko, S. Swetter, H. Blau, and S. Thrun, "Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks," Nature, vol. 542, pp. 115–118, 2017.
[4] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794.
[5] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
[6] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed. San Francisco, CA, USA: Morgan Kaufmann, 2012.
[7] World Health Organization, "Global Health Estimates: Leading Causes of Death," WHO Report, 2022.