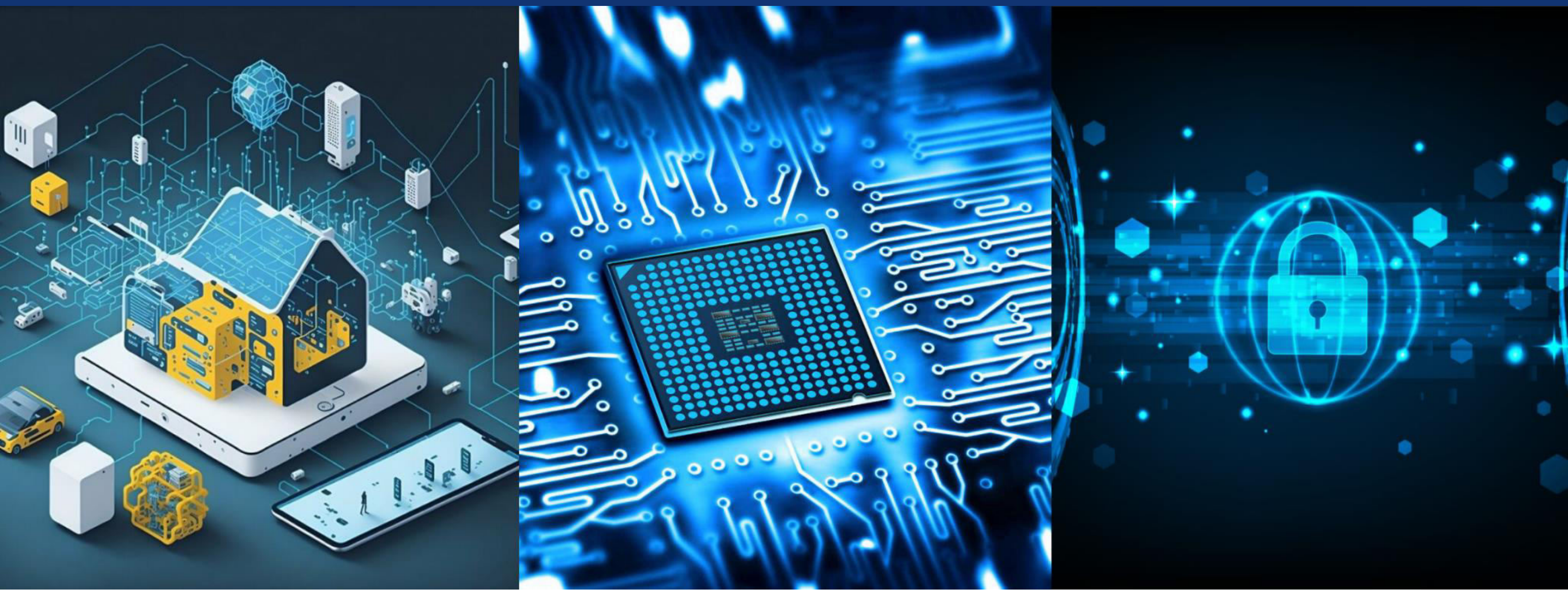
IJIRCCE

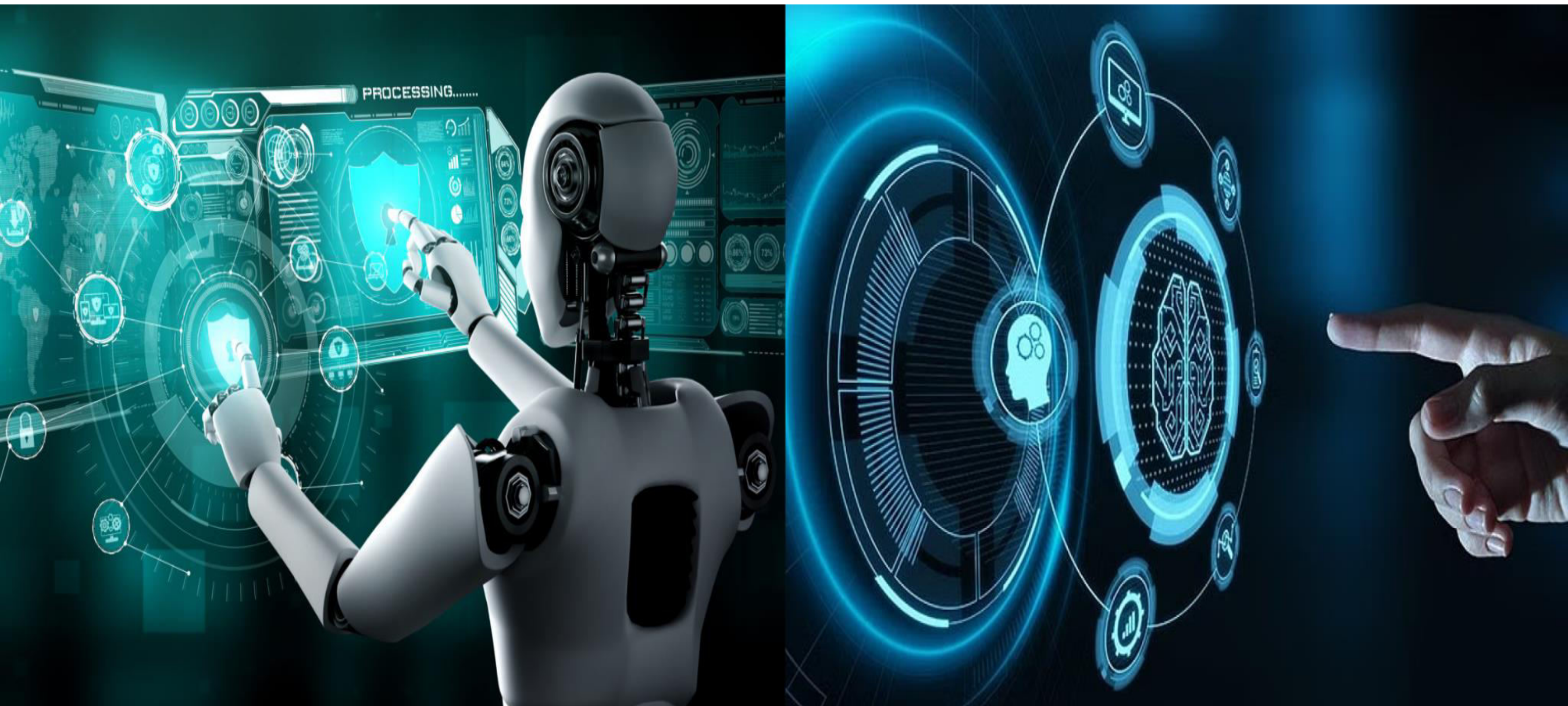
ISSN(O): 2320-9801

ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.771

Volume 13, Issue 3, March 2025

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Early Prediction of Chronic Diseases

Abhinandan R, Ashwaneeth HS, Akash G M, Dr Sheelavathy KV

B. Tech Student, Department of CSE, Reva University, Bangalore, India

B. Tech Student, Department of CSE, Reva University, Bangalore, India

B. Tech Student, Department of CSE, Reva University, Bangalore, India

Professor & Project Guide, Department of CSE, Reva University, Bangalore, India

ABSTRACT: Chronic diseases such as heart disease, diabetes, and Parkinson's disease represent a significant global health burden. These conditions often lead to severe health complications if not detected early, increasing mortality rates and healthcare costs. This paper introduces an integrated machine learning-based system designed to predict the presence of these diseases using patient-specific data. Leveraging the power of logistic regression and support vector machine (SVM) models, the system combines predictive models for heart disease, diabetes, and Parkinson's disease into a unified platform. By employing standardized datasets and implementing real-time prediction capabilities, the proposed system enhances diagnostic efficiency and accessibility. Developed with Streamlit, the web-based interface ensures ease of use for medical professionals and individuals. The integration of multiple disease prediction models into a single platform represents a breakthrough in healthcare technology, addressing scalability, user-friendliness, and real-world applicability.

KEYWORDS: chronic diseases, heart disease prediction, diabetes detection, Parkinson’s disease prediction, machine learning models, real-time disease prediction, unified prediction platform, standardized datasets, healthcare technology, scalable prediction models.

I. INTRODUCTION

Chronic diseases are the leading cause of death globally, driven by unhealthy lifestyles, aging populations, and genetic predispositions. With rising healthcare costs and an increasing number of patients requiring medical attention, early diagnosis and preventive care have become essential. Despite significant advancements in medical imaging, diagnostic tools, and data analysis, existing healthcare systems struggle with accessibility, cost-effectiveness, and scalability.

Machine learning (ML) has emerged as a promising solution for addressing these challenges. By analysing patterns in patient data, ML algorithms can provide predictive insights, enabling early interventions and personalized care. However, most existing ML-based systems focus on predicting individual diseases, such as heart disease, diabetes, or Parkinson's disease, often requiring separate platforms and lacking real-time application capabilities.

This paper presents a unified predictive system that integrates models for three critical chronic diseases into a single web-based platform. By leveraging the predictive power of ML algorithms and user-friendly technologies like Streamlit, the system is designed to deliver real-time, accurate, and scalable solutions for healthcare providers and patients.

This integrated approach not only reduces the complexity of managing multiple diagnostic tools but also improves accessibility by offering a cohesive platform that can be easily deployed and used in diverse healthcare settings. The system utilizes standardized datasets and robust preprocessing techniques to ensure reliable predictions, enhancing trust and usability for medical professionals.

© 2025 IJIRCCE | Volume 13, Issue 3, March 2025| DOI: 10.15680/IJIRCCE.2025.1303131

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

II. RELATED WORK

Numerous researchers have contributed to the development of machine learning models for disease prediction:

A. Heart Disease Prediction

Studies such as those by Narendra Mohan et al. (2021) used supervised machine learning algorithms like logistic regression, achieving an accuracy of 90.2%. However, these systems were focused solely on heart disease and did not explore integration with other predictive systems or real-time applications.

B. Diabetes Prediction

Neetu Mittal et al. (2023) developed a decision-tree- based diabetes prediction model with a 92% accuracy rate. While effective for diabetes diagnosis, the system's scope was limited to single-disease prediction without scalability for other chronic conditions.

C. Parkinson’s Disease Prediction

Yana Teletska et al. (2023) explored mathematical models and machine learning techniques for predicting Parkinson's disease progression. Although effective, their approach did not consider integration with other diseases or deployment in a user-friendly web application.

Our project builds upon these individual efforts by consolidating multiple disease prediction systems into a unified, real-time platform. This innovation addresses the limitations of standalone models and paves the way for scalable, multi-disease diagnostic solutions.

III. METHODOLOGY

To develop an accurate and effective disease prediction system, the project utilizes three distinct datasets that are specifically tailored to the diseases of interest: heart disease, diabetes, and Parkinson's disease. These datasets are sourced from well-established platforms and contain a variety of features that play a crucial role in identifying risk factors associated with each disease. The datasets vary in size and complexity, reflecting the diversity in disease characteristics and the types of patient data available.

Preprocessing steps such as data cleaning, feature scaling, and splitting into training and testing subsets were essential to prepare the data for model training. The goal was to ensure that each model could be trained on high-quality data, providing reliable predictions for real-world application. Additionally, the integration of these models into a unified web application streamlines the prediction process and ensures that predictions for multiple diseases can be made from a single, user-friendly interface.

The following section provides a detailed overview of each dataset used in the project.

A. Data Description

The project utilized three datasets, each specifically tailored for a particular disease:

1. Heart Disease Dataset:
   - Source: Kaggle
   - Size: 303 records with 14 features
   - Key Features: Age, cholesterol levels, resting blood pressure, exercise-induced angina, and maximum heart rate achieved
   - Significance: These features are critical for assessing cardiovascular risk.
2. Diabetes Dataset:
   - Source: PIMA Indians Diabetes Dataset
   - Size: 768 records with 8 attributes
   - Key Features: Glucose levels, BMI, number of pregnancies, and diabetes pedigree function
   - Significance: These variables are predictors of diabetes onset.
3. Parkinson’s Disease Dataset:
   - Source: Biomedical voice dataset
   - Size: 195 records with 22 features
   - Key Features: Jitter, shimmer, harmonic-to-noise ratio
   - Significance: These features capture vocal impairments linked to Parkinson’s progression.

B. Preprocessing

The preprocessing steps ensured data quality and consistency across all datasets:

1. Data Cleaning: Missing values were handled using mean imputation for numerical attributes and median imputation for skewed distributions.
2. Feature Scaling: Features were standardized using the StandardScaler to normalize their range, ensuring compatibility across different machine learning algorithms.
3. Train-Test Split: Each dataset was divided into training (80%) and testing (20%) subsets to evaluate model performance effectively.

C. Machine Learning Models

The models were selected and optimized based on the characteristics of each dataset:

Heart Disease Prediction:

Logistic Regression was used due to its interpretability and ability to handle binary classification tasks effectively. The model achieved an 85% training accuracy and an 82% test accuracy, demonstrating reliable performance in classifying heart disease cases.

Diabetes Prediction:

A Support Vector Machine (SVM) with a linear kernel was employed, chosen for its effectiveness in handling smaller datasets and its robustness to overfitting.

It achieved 78% training accuracy and 77% test accuracy, offering performance.

Parkinson's Disease Detection Prediction:

An SVM classifier with a linear kernel was implemented, as it performs well in high dimensional feature spaces. The model achieved 88% training accuracy and 87% test accuracy, making it highly effective for detecting Parkinson's disease.

D. Integration

The three predictive models were integrated into a unified web application using Streamlit:

User Interface: A single-page layout allows users to input patient data, such as age, glucose levels, or vocal measurements, for real-time predictions.

Backend Integration: Each model is loaded and deployed using Python libraries such as Scikit learn, ensuring efficient computation.

Real-Time Prediction: The application provides instant results for all three diseases, streamlining the diagnostic process for medical professionals.

IV. RESULTS AND DISCUSSION

The results of this integrated disease prediction system demonstrate promising accuracy across all three models. The heart disease model achieved an 85% training and 82% test accuracy, the diabetes model performed with 78% training and 77% test accuracy, and the Parkinson's disease model showed 88% training and 87% test accuracy.

These results highlight the effectiveness of each individual model. Furthermore, the integrated platform offers key advantages such as seamless scalability, real-time predictions, and a user-friendly interface. By consolidating multiple disease predictions into one system, the platform reduces redundancy, streamlines workflow, and provides consistent, reliable outcomes for healthcare professionals.

A. Heart Disease : 85% training and 82% test accuracy
B. Diabetes Disease : 78% training and 77% test accuracy
C. Parkinson's Disease : 88% training and 87% test accuracy

Comparison: Unlike standalone systems, the integrated platform ensures seamless scalability, real-time predictions, and user-friendly deployment.

Advantages:

- Consolidation of multiple diseases into one platform reduces redundancy.
- Real-time capabilities provide instant predictions, improving workflow efficiency.
- The use of standardized datasets ensures robust and reliable results.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The integrated system demonstrated high accuracy across all models, comparable to or exceeding existing systems. The unified platform minimizes redundancy and enhances user experience.

The proposed system not only excels in prediction accuracy but also stands out for its practicality and versatility in real-world applications. By integrating models for heart disease, diabetes, and Parkinson's disease into a single, cohesive platform, it simplifies the diagnostic process, reducing the need for multiple systems and associated costs. Its real-time prediction capability supports immediate decision-making, making it particularly beneficial in time sensitive clinical scenarios. Additionally, the platform’s user-friendly interface, developed using Streamlit, ensures ease of use for both medical professionals and patients. This holistic approach enhances operational efficiency while providing a scalable solution that can be expanded to include additional diseases in the future, further demonstrating the system's potential to transform preventive healthcare.

V. REAL-WORLD APPLICATIONS

The integrated disease prediction system can significantly improve healthcare delivery by enabling remote healthcare, making it easier for healthcare providers to diagnose and monitor chronic diseases without requiring in-person visits, especially in remote areas. The system also promotes preventive care, allowing for early identification of at-risk individuals and preventing disease progression. It has the potential to revolutionize community health programs by facilitating mass screenings and aiding public health decision- making. The system can be seamlessly integrated into clinical workflows, offering real-time predictions that can support decision-making during patient consultations. Ultimately, it supports healthcare's shift towards more personalized and efficient care.

A. Remote Healthcare The platform enables healthcare providers to remotely diagnose chronic diseases by collecting and analysing patient data from any location. Beneficial for rural and underserved areas where access to medical facilities is limited. Reduces the need for in-person visits, saving time and resources for both patients and clinicians
B. Preventive Care Early identification of at-risk individuals allows for timely lifestyle changes and interventions to prevent disease progression. Help reduce the burden on healthcare systems by minimizing advanced-stage disease cases.
C. Community Health Programs Ideal for deployment in community health programs and mass screening initiatives. Supports decision-making in public health policies by identifying trends and high-risk demographics.

Integration in Clinical Workflows Enhance clinical decision support systems by providing real-time disease predictions during patient consultations. Can be integrated into hospital electronic health record (EHR) systems for streamlined diagnostic workflows.

VI. SECURITY AND PRIVACY CONSIDERATIONS

A. Data Security Ensure secure storage and transmission of sensitive medical data using advanced encryption methods (e.g., AES 256 encryption). Secure access control mechanisms are implemented to restrict unauthorized access.
B. Privacy Compliance Adheres to regulatory frameworks such as HIPAA and GDPR to ensure ethical handling of patient data. Implement robust privacy policies for data collection, processing, and retention.
C. Anonymization Techniques Data used for training models is anonymized to eliminate personally identifiable information (PII). Guarantees patient privacy during both prediction and training phases.
D. User Consent and Transparency Provides a consent mechanism to inform users about how their data will be used. Offers clear, accessible terms regarding data privacy policies on the web platform.

VII. FUTURE SCOPE AND EXPANSION

A. Integration with Wearable Devices Ensure secure storage and transmission of sensitive medical data using advanced encryption methods (e.g., AES- Develop APIs to integrate data from wearable devices like smartwatches and health trackers for continuous monitoring. Real-time collection of metrics such as heart rate, glucose levels, and movement patterns to improve prediction accuracy.

B. Inclusion of Additional Diseases Expand the platform to include predictions for other chronic conditions such as stroke, Alzheimer’s, chronic kidney disease, and hypertension. Incorporate new datasets to support multi-disease

prediction models.
C. Explainable AI (XAI) Implement tools to enhance model interpretability, providing healthcare professionals with clear explanations for predictions. Enables informed decision-making, fostering trust in the system.
D. Personalized Recommendations Add modules to provide actionable insights and lifestyle recommendations tailored to individual risk profiles. Leverage historical data for personalized predictions and follow-ups.
E. Cloud and Mobile App Deployment Expand the platform to mobile applications for accessibility on smartphones. Utilize cloud computing for scalable deployment, ensuring robust performance with multiple concurrent users.

VIII. CONCLUSION

This study introduces an innovative, integrated approach to predicting three critical chronic diseases—heart disease, diabetes, and Parkinson's disease—using machine learning. By consolidating three separate predictive models into a unified, web-based platform, this system significantly enhances diagnostic efficiency, scalability,

and usability. This integration addresses key challenges in the healthcare industry, where traditional standalone systems often lack the ability to scale or provide real-time predictions.

Leveraging standardized datasets and robust preprocessing techniques, the system ensures high reliability across different healthcare settings, making it a practical tool for both clinical and non-clinical environments. The seamless combination of these models into one platform not only reduces redundancy but also provides healthcare professionals with a more streamlined and accurate diagnostic process. Furthermore, real-time functionalities, such as instant disease predictions, enable timely interventions, which are crucial for managing chronic conditions effectively.

The future potential of this project is vast, with several directions for expansion. Future improvements will include extending the platform to cover additional chronic diseases, such as stroke, Alzheimer's, and hypertension, by integrating more diverse datasets.

Another key enhancement is the incorporation of explainable AI (XAI) techniques to improve model interpretability, thereby providing healthcare professionals with transparent reasoning behind predictions. This is especially important in clinical decision-making, where understanding the rationale behind predictions can foster trust and enable informed decisions. The integration of diverse patient demographics will further enhance the model's generalizability and effectiveness.

REFERENCES

1. Narendra Mohan, Vinod Jain, Gauranshi Agrawal, "Heart Disease Prediction Using Supervised Machine Learning Algorithms," 2021.
2. Neetu Mittal, Hemangi Sharma, "Automated Disease Prediction Using Machine Learning Technology," 2023.
3. Yana Teletska et al., "Machine Learning Methods for Predicting Parkinson's Disease Progression," 2023.
4. Dr. Vibha M B, Sneha S R et al., "Exploratory Data Analysis of Heart Disease Prediction using Machine Learning Techniques-RS Algorithm," 2024.
5. S. Gupta, K. Rao, "Heart Disease Prediction Using Machine Learning Algorithms with Self-Measurable Physical Condition Indicators," SCIRP, 2023.
6. S. Mishra et al., "Heart Diseases Prediction using Machine Learning," IEEE Xplore, 2024.
7. J. Wang et al., "Application of Machine Learning in Predicting Diabetes Risk," ResearchGate, 2023.
8. R. Singh, A. Kumar, "Parkinson's Disease Detection Using Deep Learning Techniques," Elsevier, 2023.
9. T. Zhang, L. Zhao, "Early Diagnosis of Diabetes Mellitus Using Machine Learning," Springer, 2022.
10. A. Banerjee et al., "Predictive Modeling for Chronic Disease Detection Using ML Algorithms," Hindawi, 2023.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com