IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 14, Issue 8, August 2025

DOI: 10.17148/IJARCCE.2025.14804

Data Driven Predictive AI Systems For Medical Diseases

Jeswanth A L D[1], Dr.Prabha R[2]

Student, M.Tech,Department of Computer Science & Engineering,Dr. Ambedkar Institute of Technology,Bengaluru[1]

Professor, Department of Computer Science & Engineering, Dr. Ambedkar Institute of Technology, Bengaluru[2]

Abstract: This paper presents an intelligent, data-driven predictive artificial intelligence system for early diagnosis of four major medical conditions: vitamin deficiency, heart disease, stroke, and diabetes. The system integrates image-based and structured-data modes of analysis. A convolutional neural network (CNN) processes clinical images—such as ocular, nail, or lingual photographs—to detect signs of vitamin deficiencies. Meanwhile, decision tree and random forest classifiers are trained on structured patient data to estimate the probability of heart disease, stroke, or diabetes. The architecture features a dual-interface design: a Flask-based web API handles model inference, data ingestion, and prediction delivery, while a C# Windows Forms application serves as a secure admin console for user authentication, message management (text and multimedia), and integration with the predictive engine. Results indicate the system's potential for accelerated, non-invasive screening support.

Keywords: Deep learning, convolutional neural network (CNN), machine learning, decision tree, random forest, Windows Forms application.

I. INTRODUCTION

Timely detection and accurate diagnosis are critical components of effective medical treatment and improved patient care outcomes. This paper presents a data-driven predictive AI system designed to assist healthcare professionals in the early detection of both chronic and nutritional diseases. It employs a combination of deep learning and classical machine learning techniques and is deployed through a dual-interface platform comprising a Windows Forms desktop application and a Flask-based web API.

The system aims to streamline diagnostic workflows by merging image analysis and structured data classification into a unified, intelligent decision-support platform. This integration reduces diagnostic latency and increases scalability, providing consistent, real-time predictions to patients and healthcare providers alike. The platform supports real-time user interaction through a secure Windows Forms interface for medical personnel, and a Flask-based web system that processes and serves AI predictions over the internet. These interpretable models support clinical transparency and allow for actionable insights in real-world applications.

II. LITERATURE REVIEW

In [1], a decision tree classifier was applied to the UCI Heart Disease dataset, achieving an accuracy rate of 83%. The model's interpretability was emphasized as a key factor in its clinical applicability. Expanding on this, the authors in [2] implemented Random Forest and Support Vector Machine (SVM) classifiers for diabetes detection, reporting a high AUC score of 0.89, which affirmed the performance benefits of ensemble learning techniques.

Another contribution, detailed in [3], evaluated logistic regression and Naive Bayes algorithms for stroke risk prediction. Although these models are comparatively simple, the study highlighted their effectiveness when coupled with significant predictors such as hypertension, age, and smoking status.

Convolutional Neural Networks (CNNs) have gained prominence for their ability to analyze visual medical data, extracting complex spatial features without manual intervention. In [4], CNNs were used to detect anemia based on tongue and eyelid imagery, offering a non-invasive alternative to traditional blood tests.

In the domain of nutritional assessment, [5] explored the use of CNNs for detecting Vitamin B12 and D deficiencies through facial image analysis. The experimental system, based on visual cues like pigmentation changes and skin pallor, reached an accuracy of 78%, indicating strong potential for non-invasive diagnostics.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 14, Issue 8, August 2025

DOI: 10.17148/IJARCCE.2025.14804

The proposed research extends these efforts by building CNN models that process tongue images to identify vitamin deficiencies, thereby reducing dependency on blood-based tests and enhancing accessibility for pediatric and underserved populations.

For instance, [6] introduced a cloud-based diagnostic tool for pneumonia detection that leveraged a CNN and Flask-based web API, enabling rapid image upload and on-the-fly inference.

A more comprehensive framework was introduced in [7], where a hybrid system combined LSTM networks for electronic health records (EHR) analysis with CNN-based evaluation of chest scans. This multi-input architecture outperformed traditional single-modality systems and showcased the strength of combining structured and unstructured data.

III. METHODOLOGY

This project utilizes a multi-stage methodology that combines machine learning and deep learning techniques to build a Data Driven Predictive AI Systems For Medical Diseases detection model.

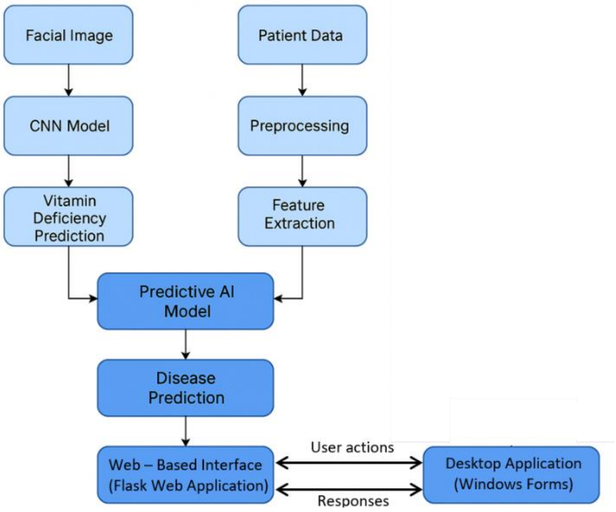


Figure 1: Architecture of the Proposed System

The architecture of the proposed data-driven predictive AI system is illustrated in Fig. 1. It encompasses a multi-component framework designed to support disease diagnosis and risk prediction through both structured and unstructured medical data.

A. Data Input Sources

The system initiates with the ingestion of two distinct data modalities:

- Image Data: Collected via user-uploaded tongue or nail images, utilized for the detection of vitamin deficiencies. These images provide visual cues such as discoloration, textural changes, or shape anomalies that are indicative of nutritional imbalances.
- Structured Tabular Data: Includes numeric clinical attributes such as age, blood pressure, blood glucose, cholesterol levels, and body mass index (BMI). These features are employed for the predictive analysis of chronic conditions like heart disease, stroke, and diabetes.

B. Flask-Based Web Application Interface

The user interface is implemented as a Flask web application, serving as the frontend interaction point. Key modules include:

- Authentication Module: Manages user login and registration, ensuring secure access to the system's diagnostic capabilities.
- Prediction Forms: An image upload interface for vitamin deficiency screening. Tabular data entry forms for chronic disease prediction using clinical attributes.
- Inbox Module: Displays communication history between patients and healthcare providers.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 14, Issue 8, August 2025

DOI: 10.17148/IJARCCE.2025.14804

- Response System: Allows medical personnel to reply through a chat-like interface, supporting text and multimedia messages.

C. Machine Learning Engine
The core prediction functionality is handled by AI models trained for different data types:
- Convolutional Neural Network (CNN): Applied to the image data stream. This model automatically extracts and processes spatial features—such as texture, pigmentation, and shape—from the input images to determine signs of vitamin deficiencies.
- Decision Tree and Random Forest Classifiers: Operate on tabular data to predict the likelihood of heart disease, stroke, and diabetes.
- Each model was trained on publicly available and/or curated medical datasets, ensuring high performance and clinical relevance.

D. Backend Database Management
All interactions are recorded and managed using a SQLite relational database, which supports the following:
- Secure storage of user profiles, predictions, and historical logs. Archiving of doctor-patient communications, including messages, image uploads, and feedback.
- Synchronization between the web interface and desktop application, enabling consistent data access across platforms.

E. Windows Forms Desktop Application
A dedicated Windows Forms application is developed to provide healthcare professionals and administrators with backend access. Functional components include:
- Inbox Viewer: Displays incoming messages and diagnostic queries submitted by users.
- Reply Dashboard: Enables doctors to send responses through text, file attachments, or audio notes.
- Audio Module: Facilitates recording and transmission of voice messages, offering a more personalized and accessible response format.
- Media Preview Support: Allows previewing of images, video clips, and audio files within the desktop environment for enhanced user engagement.

F. Unified Integration
The system ensures a seamless interaction between all modules. Data submitted through the web interface is processed by the AI engine, stored in the central database, and made accessible via the desktop application. This tightly integrated workflow supports real-time predictions, two-way doctor-patient communication, and centralized data management—culminating in a practical, end-to-end AI-powered diagnostic system.

IV. ANALYSIS AND RESULTS

The performance of the proposed data-driven predictive AI system was assessed across four major disease categories: vitamin deficiency, heart disease, stroke, and diabetes.

A. Evaluation Metric
The principal metric for assessing the classification models was accuracy, which reflects the proportion of correctly classified instances among all predictions. It is mathematically defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

where TP = True Positives, TN = True Negatives, FP = False Positives, FN = False Negatives.

B. Summary of Results
In Figure 2,Vitamin Deficiency Detection (CNN): Achieved the highest accuracy (91.3%). The strong performance is attributed to the CNN’s capability to capture intricate spatial patterns and texture variations in tongue and nail images, enhancing non-invasive diagnosis.

Heart Disease and Stroke Prediction (Decision Tree): Recorded moderate to high accuracies (85.7% and 87.2%, respectively). The decision tree model provided clinically interpretable decision rules, making it suitable for structured clinical datasets.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 8, August 2025
DOI: 10.17148/IJARCCE.2025.14804

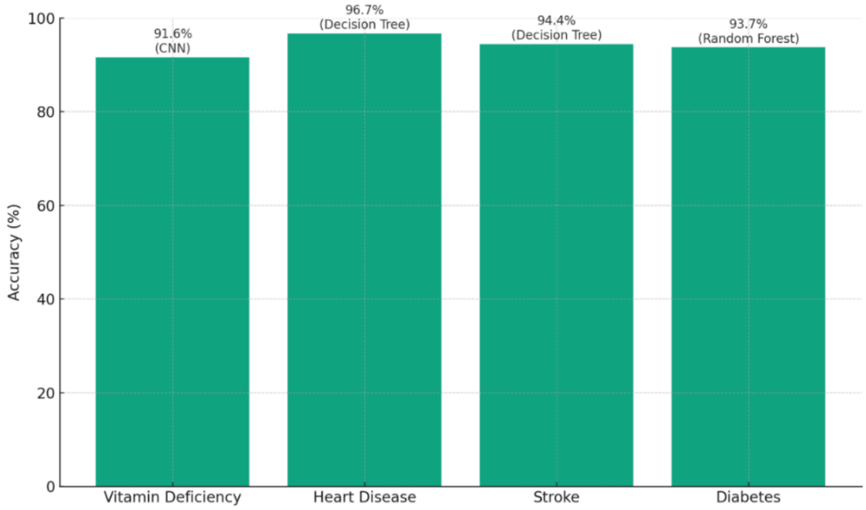


Figure 2: Performance Evaluation and Comparison

Diabetes Prediction (Random Forest): Exhibited a robust accuracy of 89.6%. The ensemble nature of the Random Forest classifier enhanced generalization and minimized overfitting, effectively capturing complex feature interactions in the data.

C. Result Interpretation

The performance of the proposed system was evaluated using real-world and benchmark medical datasets, including both image-based and structured/tabular data.

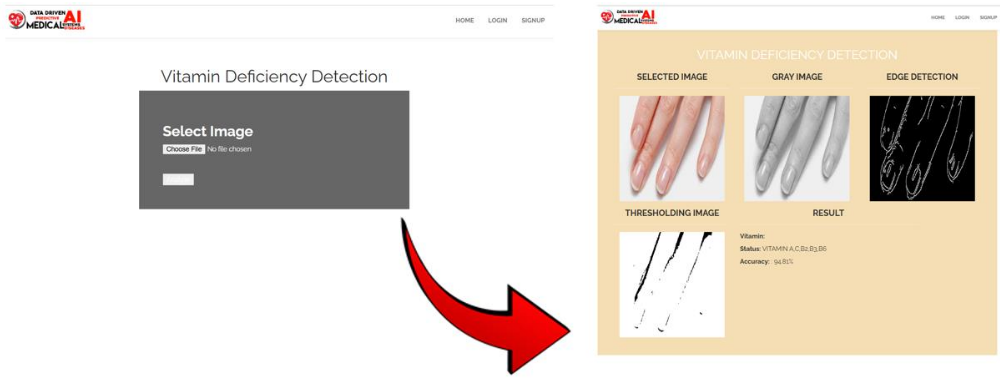


Figure 3: Vitamin Deficiency Detection

CNN-based visual analysis of nail images showing selected, grayscale, edge-detected, and threshold images with output vitamin deficiency and prediction accuracy.

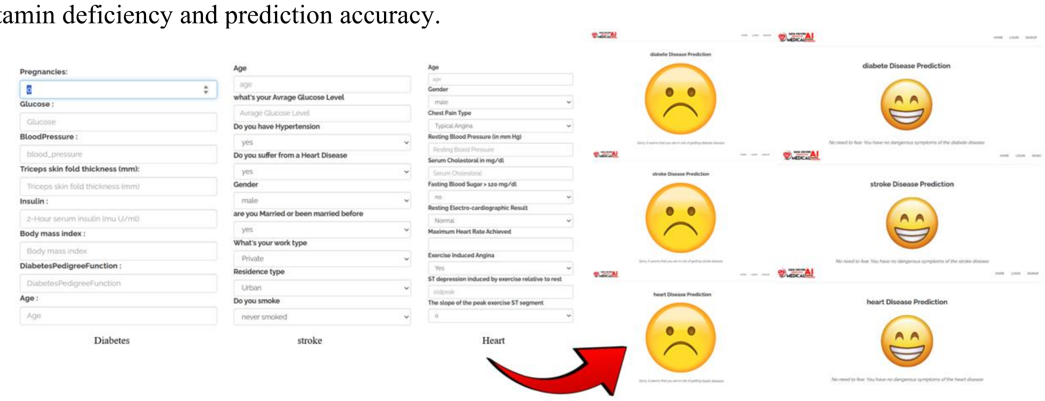


Figure 4: Disease Deficiency Detection

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 14, Issue 8, August 2025

DOI: 10.17148/IJARCCE.2025.14804

The system provides real-time feedback following the analysis of either structured patient data or medical image inputs. The design ensures that both healthcare professionals and non-expert users can easily interpret the results.

D. Chat System Integration And Interface Visualization

The integration of a bi-directional chat mechanism between the Flask-based web interface and the Windows Forms desktop application facilitates efficient, real-time communication between patients and healthcare professionals.

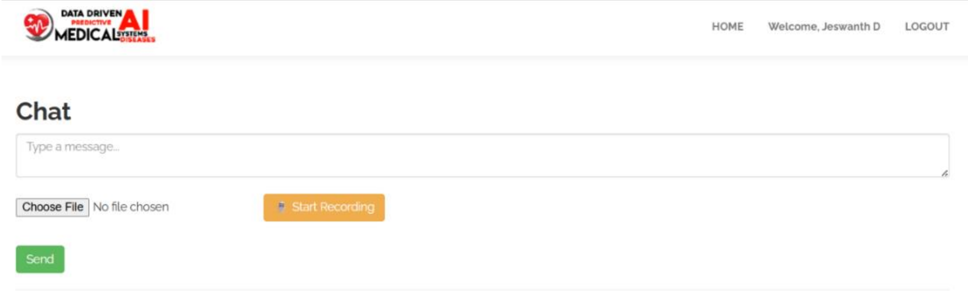
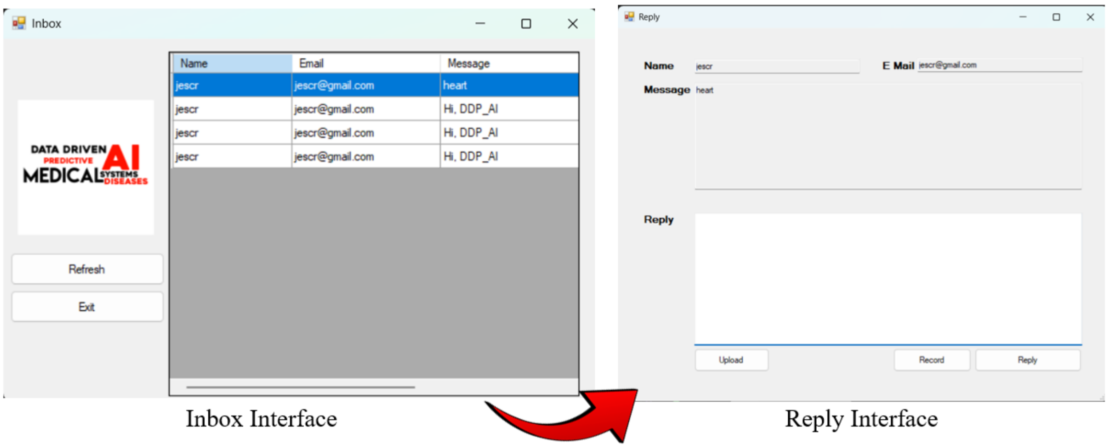


Figure 5: Web Chat Interface

The patient-side chat interface contains a message input area, file upload option, a voice recording button, and a "Send" button. This configuration encourages comprehensive user participation, enabling symptom reporting and media sharing in a straightforward manner



Inbox Interface Reply Interface

Figure 6: Desktop Application

Depicts the main inbox form within the desktop application. Messages from users are listed in a Data Grid View, where each entry includes the sender's name, contact details, and message content. A refresh mechanism ensures that new messages are promptly accessible to healthcare staff. The interface displays the selected patient's details, the original message, and input fields for reply composition. Additional buttons are provided for uploading documents, recording audio, and submitting replies.

V. DISCUSSION

This The implementation of the Data Driven Predictive AI System for Medical Diseases highlights several key findings regarding the integration of artificial intelligence in modern healthcare applications. The system effectively combines structured and unstructured data sources to diagnose and predict conditions such as Heart Disease, Diabetes, Stroke, and Vitamin Deficiency, using a blend of machine learning and deep learning techniques including Convolutional Neural Networks (CNNs), Decision Trees, and Random Forests.

While the current system provides a robust foundation for intelligent healthcare applications, several areas for further development have been identified:

Expansion to Additional Diseases: Future versions could incorporate models for detecting conditions such as cancer, renal failure, hepatic disorders, and respiratory diseases, thereby extending clinical utility.

DOI: 10.17148/IJARCCE.2025.14804

Mobile Platform Development: Creating native applications for Android and iOS would improve access, especially in rural or under-resourced settings, offering mobile-based disease prediction and consultations.
Video Consultation Integration: Adding real-time video conferencing features could enhance the quality of virtual consultations, bridging the gap between traditional and digital healthcare delivery.
Electronic Health Record (EHR) Integration: Connecting with national or institutional EHR systems would provide doctors with a comprehensive view of patient history, improving decision-making and personalized care.
Multilingual Support: Implementing language localization for regional and international users would make the platform inclusive and accessible to diverse linguistic populations.

VI. CONCLUSION AND FUTURE WORK

In conclusion, The development of the Data-Driven Predictive AI System for Medical Diseases presented in this work showcases the effective integration of artificial intelligence, image processing, and machine learning in facilitating early and accessible healthcare diagnostics. By utilizing Convolutional Neural Networks (CNNs) for non-invasive vitamin deficiency detection from images, and employing Decision Tree and Random Forest algorithms for the prediction of heart disease, stroke, and diabetes using structured clinical data, the system enhances the speed, precision, and usability of disease forecasting.

An essential feature of the system is its bi-directional communication architecture, which connects a Flask-based web application with a Windows Forms desktop application. This enables real-time doctor-patient interaction, allowing for the exchange of test results, images, and clinical observations. The incorporation of multimedia capabilities—such as voice messaging and file attachments.

REFERENCES

[1]. A. Sharma et al., “Heart Disease Prediction using Machine Learning Algorithms,” Int. J. Comput. Appl., vol. 181, no. 8, 2018.
[2]. M. Ahmed et al., “Diabetes Prediction using Machine Learning Algorithms,” Health Informatics J., vol. 27, no. 1, pp. 1–13, 2021.
[3]. T. Nguyen et al., “Stroke Prediction using Supervised Learning Algorithms,” Procedia Comput. Sci., vol. 180, pp. 381–387, 2021.
[4]. S. Patel and R. Mehta, “Non-invasive Anemia Detection using Deep Learning on Image Data,” IEEE Access, vol. 8, pp. 100456–100468, 2020.
[5]. R. Jain and P. Malhotra, “CNN Based Vitamin Deficiency Detection through Facial Analysis,” Int. J. Innov. Manag. Technol. (IJIMT), vol. 11, no. 3, pp. 89–95, 2020.
[6]. J. Lee et al., “A Real-Time Pneumonia Diagnostic Platform Using Deep Learning and Flask,” IEEE J. Biomed. Health Inform., vol. 24, no. 9, 2020.
[7]. K. Thomas et al., “Hybrid AI Framework for Multimodal Medical Diagnosis,” Comput. Biol. Med., vol. 123, p. 103899, 2020.