Archives available at journals.mriindia.com

International Journal on Advanced Computer Engineering and Communication Technology

ISSN: 2278-5140
Volume 14 Issue 01, 2025

AI-Driven Symptom Analysis and Disease Prediction Assistant for Healthcare

Soham Gholap[1], Sarthak Ghodke[2], Parth Mane[3], Vivek Vanave[4], Dr. Dipannita Mondal[5]

[1-4]U.G. Students, Department of AI/DS, Dr. D Y Patil College of Engineering and Innovation, Varale, Talegaon Dabhade, Pune, India

[5]Head of Dept. of AI/DS, Dr. D Y Patil College of Engineering and Innovation, Varale, Talegaon Dabhade, Pune, India

Peer Review Information

Submission: 21 Feb 2025
Revision: 25 March 2025
Acceptance: 30 April 2025

Keywords

Artificial Intelligence (AI)
Machine Learning (ML)
Healthcare Assistant
Symptom Analysis

Abstract

The use of Artificial Intelligence (AI) and Machine Learning (ML) is revolutionizing the healthcare domain, particularly in enhancing diagnostic processes. This paper introduces a smart assistant that employs AI and ML to interact with patients, collecting symptom data, past medical records, and image inputs. It then analyzes this information to predict possible diseases and recommend preliminary remedies. The outcomes are forwarded to healthcare professionals for final assessment, thereby reducing workload and improving efficiency. The system aims to enhance early diagnosis, ensure timely interventions, and support medical experts with advanced data-driven insights—all while keeping the human expert as the final authority.

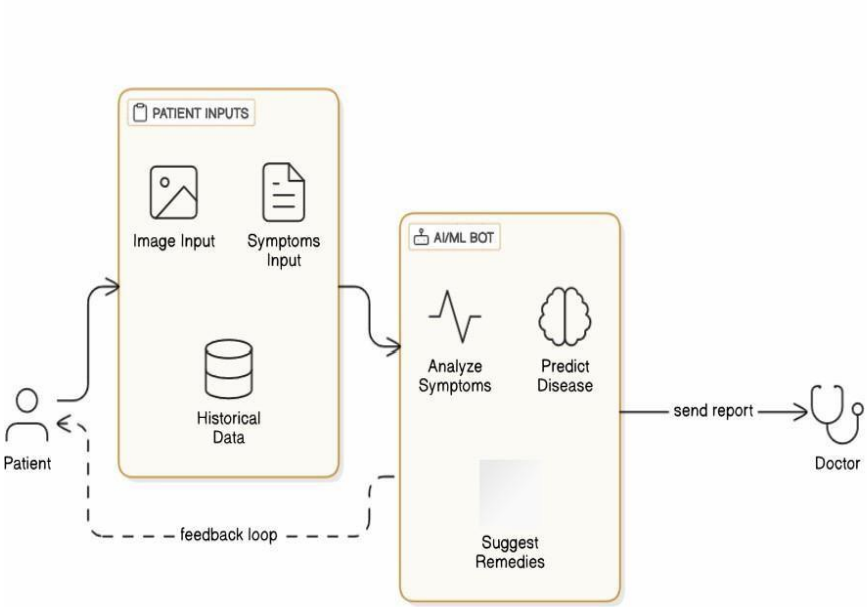
PROPOSED SYSTEM

Our project involves building an intelligent assistant that interacts with patients, gathering symptom-related information, predicting diseases, and suggesting primary treatment options. This processed data is then forwarded to doctors for their evaluation. The existing healthcare infrastructure struggles with high patient volumes and delayed diagnostics due to manual data collection. Our system mitigates these challenges by automating early-stage analysis, enabling healthcare providers to prioritize complex cases effectively. This system should be seamlessly integrated into all healthcare workflows, bringing well-rounded and accurate patient information that will enable doctors to accelerate their decision-making process to focus on highly complex cases.

Scope of the System:

1. **Assisting Doctors in Pre-Processing Data:** Reduces doctors' workload by preprocessing patient data, offering preliminary diagnoses, and providing actionable insights to help with complex cases.
2. **Improving Patient Experience:** Offers instant feedback on symptoms, suggests potential diagnoses, and advises on next steps, reducing patient anxiety and enabling proactive action.
3. **Reducing Diagnosis Time:** Speeds up diagnostic processes, especially in underserved areas, with automated symptom checks and preliminary analyses, enabling quicker interventions.
4. **Multimodal Input Handling:** Supports diverse inputs like text descriptions, images (e.g., rashes, X-rays), and historical medical data for comprehensive analysis.
5. **Scalability and Flexibility:** Adapts across medical domains, from dermatology to cardiology and general practice, ensuring versatility for diverse healthcare needs.

ARCHITECTURE DIAGRAM



ALGORITHMS

1. Support Vector Classifier (SVC)

Purpose: Finds the optimal hyperplane for classification, ideal for linearly separable data.
Why Used: Effective for high-dimensional data and to assess performance against other mode.

2. Random Forest Classifier

Purpose: Builds multiple decision trees, outputs majority vote.
Why Used: Robust to overfitting, handles mixed feature types, and provides feature importance.

3. Gradient Boosting Classifier

Purpose: Builds models sequentially, minimizing loss and learning from errors. **Why Used:** Often outperforms Random Forest in smaller datasets, reduces bias and variance.

4. k-Nearest Neighbors (KNeighbors)

Purpose: Classifies based on majority vote of nearest neighbors.
Why Used: Baseline model for comparison, effective in smaller datasets.

5. Multinomial Naive Bayes (MultinomialNB)

Purpose: Probabilistic model for classification with discrete features. **Why Used:** Simple, fast, and effective for text-based or count data.

6. Cross-Validation (cross_val_score)

Purpose: Assesses model generalization by training on multiple data splits.
Why Used: Mitigates overfitting and ensures robust accuracy comparison

MATHEMATICAL MODEL

i. Input Variables (Patient Data):

S= Set of Symptoms S={$s1$,$s2$,...,sn};
Where,
I = Image inputs (e.g., medical scans or photos)
I = {$i1$,$i2$,...,im};
H = Historical data (patient's medical history)
H={$h1$,$h2$,...,hk}

ii. Feature Extraction Function:

$F(S,I,H)$ = Function that extracts features from symptoms, images, and historical data.
$F(S,I,H)$ = {$f1$,$f2$,...,fp};
$F(S,I,H)$= {f 1 ,f 2 ,...,f p },
Where,
f if I represents a relevant feature for diagnosis.

iii. **Disease Prediction Model**:

D = Predicted Disease
$P(D|F(S,I,H))$ = Probability of a disease given the extracted features using a classifier (e.g., Decision Tree, SVM, or Neural Network).

iv. Remedy Suggestion Model:

R = Set of Remedies R={$r1$,$r2$,...,rq};
$P(R|D*)$ = Probability of a remedy given the predicted disease.
$R*$ = argmax$P(R|D*)$ $R*$ = argmaxP(R|D*), Where, $R*R*$ is the recommended remedy.

v. Doctor Feedback Loop:

The final decision is shared with the doctor $Dadmin$D for review;
$Dfinal$=$\alpha Dadmin$+(1−α)$D*$D final =αD admin +(1−α)D ∗ , Where,
$\alpha\alpha$ is a feedback factor from the doctor

RESULT

Aspect	Existing System	Proposed AI System
Symptom Collection	Manual, time-consuming	Automated via bot, quick and efficient
Diagnosis Process	Relies solely on doctor's analysis, potential for human error	AI-driven preliminary diagnosis with high accuracy
Data Utilization	High workload due to manual symptom collection and analysis	Reduced workload through preprocessed, AI-analyzed data
Response Time	Longer wait times for patients due to overload doctors	Faster response and diagnosis with AI pre- processing
Personalization	Generic diagnosis based on immediate symptoms	Personalized diagnosis using Historical and real- time data

EXPERIMENTAL RESULTS

Fig.1 Home Page

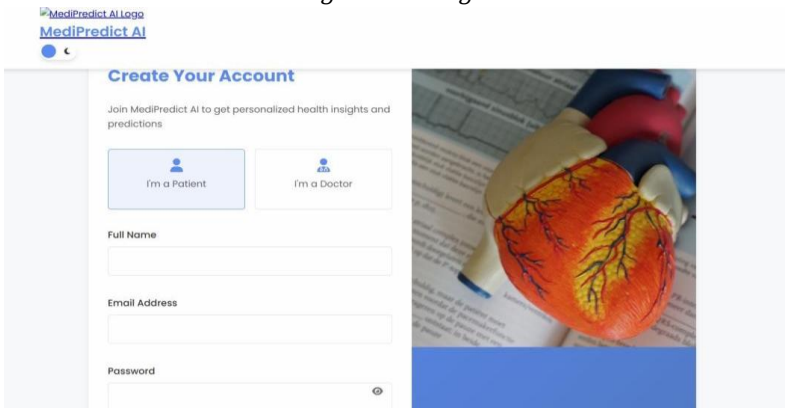


Fig.2 Sing Up Page

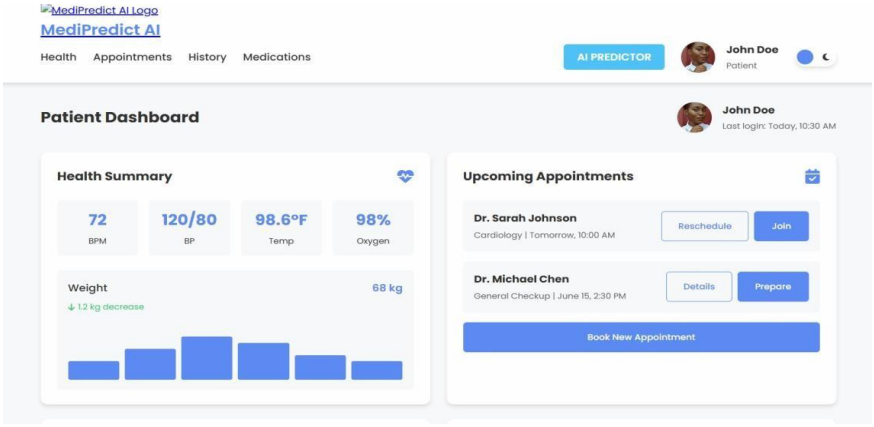


Fig.3 Patient Dashboard

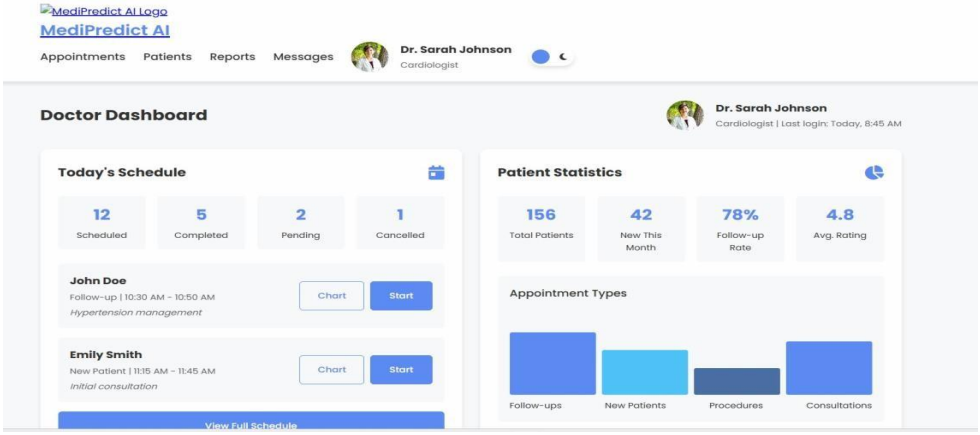


Fig.4 Doctor Dashboard

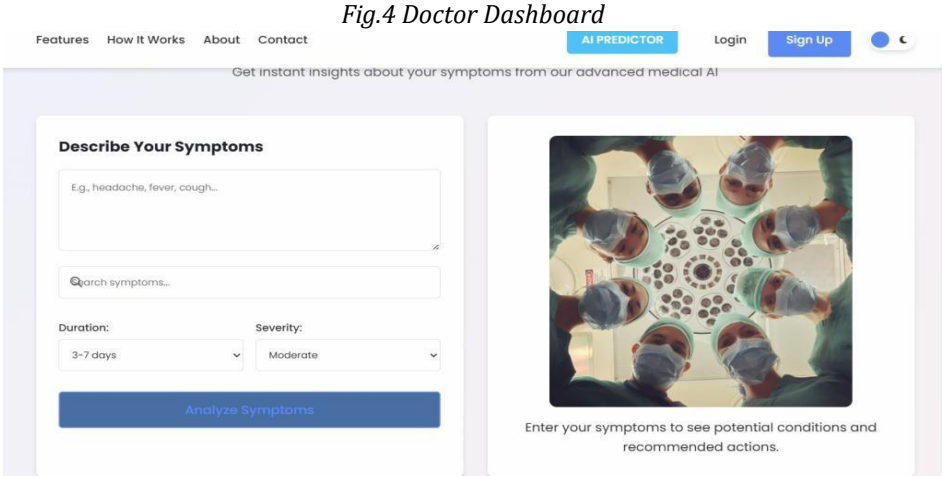


Fig.5 AI-Predictor

PURPOSE OF STUDY

The purpose behind this project stems from the growing demand for personalized and efficient healthcare services. With the increasing burden on healthcare providers and the need for quicker diagnostic processes, there is a critical need for technologies that can assist in early disease detection and patient management. By leveraging artificial intelligence and machine learning, this project aims to bridge the gap between patients and healthcare providers, ensuring that critical health data is analyzed accurately and efficiently.

APPLICATION

- Preliminary Diagnosis
- Symptom Analysis
- Personalized Healthcare

ADVANTAGES

1. **Reduced Wait Times:** Faster initial diagnosis reduces patient wait times and improves access to timely care.
2. **Improved Healthcare Efficiency:** Frees up doctor time to focus on complete case, enhancing overall healthcare service delivery.
3. **Early Disease Detection**: AI can detect potential diseases early, leading to quicker interventions and better health outcomes.
4. **Personalized Patient Care:** Tailored diagnoses and remedies based on individual data improve patient satisfaction and outcomes.

DISADVANTAGES

1. Health rlated data leakage is a concern.
2. Large amount of module training as diseases can be of various types.

CONCLUSION

The integration of AI and ML in healthcare through the proposed assistant system offers a transformative approach to diagnostics. The system reduces the burden on doctors while maintaining diagnostic accuracy and reliability by automating symptom analysis and disease prediction. This collaborative model between AI and medical professionals ensures optimal patient care. Future work will be directed to further enhancement of the systems by developing advanced neural networks, integration of real-time feedback, and perfecting the adaptability of the system to new emerging diseases. Further research shall also look into ethical implications and data privacy in safe and fair usage.