Archives available at journals.mriindia.com

International Journal on Advanced Computer Theory and Engineering

ISSN: 2319-2526
Volume 14 Issue 01, 2025

AI Healthcare Disease Prediction Intelligence System based on Symptoms Using Machine Learning

Sushma S. Deotale[1], Dr. Poonam N. Railkar[2]
*[1,*2]Smt. Kashibai Navale College Of Engineering, Vadgaon, Pune, India*
[1]sushmadeotale123@gmail.com, [2]poonamrailkar@gmail.com

Peer Review Information

Submission: 20 Jan 2025
Revision: 22 Feb 2025
Acceptance: 27 March 2025

Keywords

AI In Healthcare
Disease Prediction
Symptom Analysis
Machine Learning
Medical Diagnosis

Abstract

The Medical Recommendation System is an AI-driven web application designed to help users identify possible diseases based on their symptoms. It leverages machine learning algorithms to analyze both typed delivering predictions on likely illnesses. Additionally, the system provides comprehensive healthcare guidance, including disease overviews, precautionary steps, medication suggestions, exercise routines, and dietary plans. The development involves data collection, preprocessing, and training deep learning models, with performance evaluated through accuracy and precision metrics. For many medical organizations, disease prediction is very important for making the best possible healthcare decisions. Machine Learning is a field where we can develop a model to learn machines to make decisions on their own from real-time data and from past experience. The research involves collecting relevant medical data, training deep learning models, and evaluating their predictive capabilities through metrics such as accuracy and precision.

Introduction

Artificial intelligence is changing healthcare by making it more accurate, efficient, and personalized. One key advancement is the Drug Recommendation System, which helps doctors and patients choose the right medicines based on the diagnosis, medical history, and symptoms. Using machine learning and natural language processing, it ensures better and safer treatment. Traditional prescribing depends on a doctor's judgment, which can sometimes lead to mistakes. AI improves this by suggesting more reliable options, especially in areas where expert medical advice might be limited. This project looks into how such a system is built using drug interaction databases and patient data. The goal is to support better decisions, improve patient care, and reduce harmful side effects.[1]Artificial Intelligence (AI) has revolutionized the healthcare industry by enhancing the accuracy, efficiency, and personalization of medical treatments. A Drug Recommendation System, powered by AI, serves as a vital tool for both healthcare providers and patients, offering medication suggestions based on diagnostic data, patient histories, and presenting symptoms. By leveraging machine learning and natural language

processing, such systems improve treatment safety and efficacy. Unlike traditional prescribing methods, which rely heavily on clinician intuition and may lead to errors, AI-driven solutions reduce risk and optimize care—especially in regions with limited medical resources. This research explores the creation and implementation of a system that integrates drug interaction databases with patient-specific data to support more effective clinical decision-making, ultimately improving outcomes and minimizing adverse effects.[4]

This project focuses on the design and implementation of a drug recommendation model

© 2025 The Authors. Published by MRI INDIA.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

that incorporates drug interaction databases and individual patient profiles. The solution enhances therapeutic accuracy and reduces adverse effects, especially in under-resourced areas, by augmenting clinical decisions with intelligent, data-driven support.

Our AI-driven system leverages Natural Language Processing (NLP) and cosine similarity algorithms to identify and suggest alternative medications based on their chemical and therapeutic properties.

- Recommends AI-generated alternative medicines.
- Employs NLP & cosine similarity for precise drug matching.
- Finds medications with similar active ingredients.

Literature Review

Mayur Gadekar, [1] Via assisting doctors in early clinical diagnosis and prediction, a classification method that was constructed employing machine learning techniques was intended to greatly aid inside this settlement for mental well-being challenges. One representative collection containing 49201 patient records having diagnosis covering 40 diseases were selected for the study. The multiple regression was made up of 42 disorders. There must have been 95 of 132 independent factors (symptomatology) that were clearly related to diseases. In just this research study, a sickness estimation system was originally developed using machine learning algorithms like Classifier, Random Forest classifier, and the Naive Bayes classifier was exhibited. In just this study, 3 systems' effectiveness on a health file is thoroughly compared, with each algorithm producing results with just a reliability of upwards to 94%.

Tania Nipa [2] Through examining key metrics, thes research seeks to find trends across different based On supervised model types towards sickness detection. Regarding classifiers, the KNearest Neighbor, Decision Tree (DT), and Naïve Bayes classifier (NB) algorithms attracted the most interest (KNN). Svm Classifier (SVM) is indeed the best at detecting parkinson's, as according to studies. Using Logistic Regression (LR) performed extremely well enough in terms of predicting heart conditions. Lastly, recommendations were performed employing Convolutional Neural Networks (CNN) and Random Forest (RF), accordingly, both precise specific diseases of infectious symptoms.

Mamata Garanayak [3] About this research, researchers propose a hybrid machine learning framework composed of support vector machines and evolutionary algorithms. Data sets for the liver, diabetes, and heart, that everything retrieved either from a reputed university were utilized to assess our algorithm. On the data sets, researchers evaluated their algorithms. These were 75.4% effective on the Cleveland Heart dataset. By maintaining all of the parameters, researchers were able to cut the feature count from Thirteen to Ten with sustaining overall compromise on accuracy as evaluated by the SVM alone across the entire data. For such liver datasets, they achieved the highest precision (78.6%). By minimizing the number of components, they were able to achieve marginally lower accurateness, however, they were both still well within acceptable parameters.

K. Gaurav, [4] This test measures the patient's signs as input and determined the risk that the sickness may arise. Utilizing a decision-tree classifier, disease prediction is actually achieved. The frequency of an illness is computed by a decision tree classifier. Appropriate diagnosis data management improves in earlier disease classification and clinical management as big data demand increases inside the healthcare and biomedical industries.

Keniya, R., [5] it created a methodology to predict diseases employing multiple Machine learning techniques. diseases were included in the dataset that has been examined.

This assessment system includes an outcome as such sickness which a person might well be encountering in light of the symptoms, age, and gender of the patient. Through comparing the different algorithms, a balanced KNN algorithm generated the greatest outcomes. This balanced KNN algorithm had a predictive performance around 93.6%. If an illness is identified earlier, our diagnostics program may play the role of a doctor, enabling appropriate care as well as the possibility of ensuring safety. designations.

Methodology

Our system integrates various AI, machine learning, and NLP techniques to provide accurate and explainable medical recommendations. The development process involves the following core components:

- Data Collection & Preprocessing
- Dataset Description & Accessibility
- Feature Engineering & Similarity Matching
- Model Development
- Recommendation Engine
- Web Application Interface
- Symptoms Dataset

1. **Data Collection & Preprocessing:**
   Medical datasets were cleaned and structured using Pandas and NumPy. Data was serialized and stored efficiently using Pickle for streamlined access during training and inference.
2. **Dataset Description & Accessibility:**
   The ensures the confidentiality and protection of personal health information by strictly prohibiting the unauthorized disclosure of a patient's medical records without their explicit consent. Access to

government-maintained health records and medical databases typically requires multiple levels of authorization and licensing. Due to these legal and ethical restrictions, our research relies on publicly available datasets that are freely accessible and downloadable online.

3. **Feature Engineering & Similarity Matching:**
Extracted drug-related features and computed similarity scores using Natural Language Processing (NLP) and Cosine Similarity. Employed Hugging Face Transformers to generate contextual embeddings for drug descriptions and ingredients.

4. **Model Development:**
The development of the predictive model involved several stages, beginning with data preprocessing and feature selection. The raw dataset, derived from publicly available medical records and curated symptom-drug associations, was cleaned and transformed to ensure consistency and quality.
The AI Healthcare Disease Prediction System based on Symptoms Using Machine Learning typically uses the following machine learning models:

- Decision Tree
  Simple and interpretable. Works well with symptom-based classification.
- K-Nearest Neighbors (KNN)
  Classifies based on similarity to known cases. Easy to implement, good for small datasets.
- Naive Bayes
  Based on probability and Bayes' Theorem. Performs well with categorical symptom data.
- Random Forest
  Ensemble of decision trees. High accuracy and handles overfitting better than a single decision tree

Fig 1: Comparison of Model Performance Metrics

Here's a bar graph showing sample values for Accuracy (0.88), Precision (0.85), and Recall (0.80)—common metrics used to evaluate machine learning models as shown in Figure 1.

5. **Recommendation Engine:**
Developed an AI-powered engine to suggest alternative medications based on similarity in therapeutic properties and composition. Leveraged cosine similarity to match drug vectors derived from NLP embeddings.

6. **Web Application Interface:**
Built an intuitive front-end using Streamlit, allowing users to input symptoms or drug names and receive recommendations in real-time.

7. **Symptoms Dataset:**
To prepare this dataset for analysis, data cleaning was necessary. The "Count" column was removed because it was irrelevant to the study. Null values in the "Disease" column were addressed using the drop function. Both the "Disease" and "Symptom" columns were cleaned to retain only the names, removing any unrelated data as shown in Figure 2.

	Disease	Symptom
0	hypertensive disease	[pain chest, shortness of breath, dizziness, a...
1	diabetes	[polyuria, polydypsia, shortness of breath, pa...
2	depression mental	[feeling suicidal, suicidal, hallucinations au...
3	depressive disorder	[feeling suicidal, suicidal, hallucinations au...
4	coronary arteriosclerosis	[pain chest, angina pectoris, shortness of bre...

Fig 2: Symptoms dataset after pre-processing

Proposed System

The AI Medicine Recommendation System evaluates a user's ability to self-diagnose a range of pre-defined diseases based on reported symptoms. Following this self-assessment, the system recommends appropriate medications, safety precautions, exercise routines, and

specialized diet plans that align with the user's assessment grade.
To enhance user experience, the platform supports multiple input methods including voice recognition, keyboard entry, and other interfaces.

It matches user-reported symptoms with trained machine learning models using the latest medical datasets, providing accurate disease predictions and actionable health recommendations.
The system utilizes a trained model fine-tuned using supervised learning techniques such as Decision Trees, Random Forests, and Neural Networks. Integrated with AI, it not only predicts the disease but also provides concise descriptions, preventive strategies, medication recommendations, fitness routines, and dietary guidance.
The user interface is designed to be responsive and intuitive, with built-in search functionality. The system is scalable and continuously improves its prediction accuracy with new data. Additionally, it integrates with Electronic Health Records (EHR), databases, and other diagnostic tools to ensure transparency, reliability, and an enhanced user experience.
This is done using the trained model which is further fine-tuned with the Decision Trees, Random Forest, or Neural Networks supervised learning techniques. The AI integrated recommends and describes the disease in brief along with preventive measures, medications, suitable workouts, and dietary measures. The comprehensive responsive UI of the system designed is user-friendly and enables search.
These measures are scalable to increases the prediction accuracy of the model with new data input. As a prognostic tool, the AI system has EHR systems, databases, and many other tools integrated into it ensuring the user has always set expectations regarding the system while enhancing their experience and satisfaction with it.

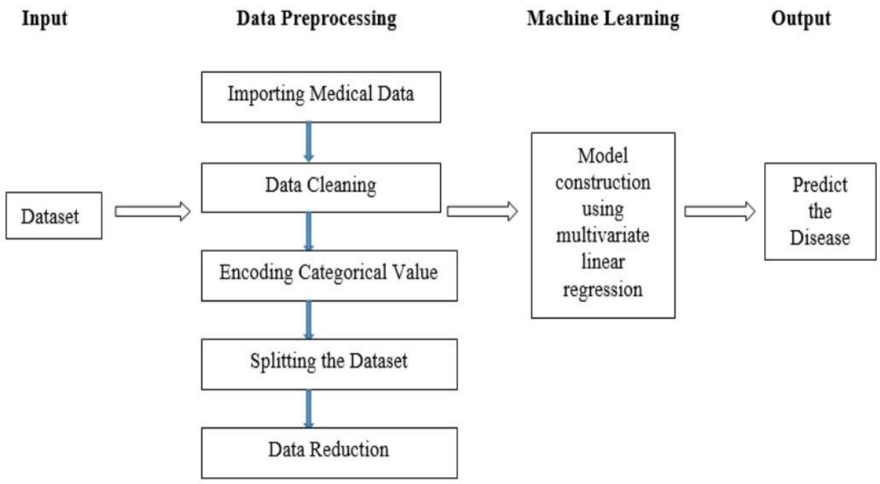

Fig 3: System Architecture

The architecture of a medical disease prediction system using machine learning begins with importing a medical dataset**,** which may include patient demographics, clinical measurements, and diagnostic test results as shown in Figure 3.

Result And Discussion.

The Drug Recommendation System employs machine learning to suggest optimal medications based on a patient's diagnosis, symptoms, and medical history. By using models like decision trees, random forests, and deep learning, it enhances prescription accuracy. The system improves healthcare outcomes by reducing errors and supporting personalized treatment. Integration with electronic health records (EHRs) further aligns the system with precision medicine standards.
Using machine learning, the Drug Recommendation System picks the best medicines based on the patient's health history, diagnosis, and symptoms. It uses smart algorithms like decision trees and deep learning to make fewer mistakes and improve treatment. By connecting with health record systems, it gives more personalized and effective care.
The Drug Recommendation System leverages advanced machine learning methodologies—including decision trees, random forests, and deep learning architectures—to identify optimal pharmacological interventions tailored to individual patients' medical history, diagnosis, and symptom profiles.
The system significantly improves therapeutic precision and minimizes prescription errors. Furthermore, its integration with electronic health information systems enhances alignment with contemporary precision medicine protocols.

Fig 4: Disease Prediction

The goal is to predict the most probable disease(s) based on symptoms input by a user, using machine learning models trained on medical datasets as shown in Figure 4.

Fig 5: Disease Prediction Based on Symptoms

This system uses machine learning (ML) to predict possible diseases by analyzing the symptoms entered by a user (patient or doctor). It's an AI-powered decision support system designed to help with early diagnosis and medical guidance as shown in Figure 5.

Fig 6: Drug Recommendation System

A Drug Recommendation System uses AI and machine learning to suggest appropriate medications or treatments after predicting a disease based on patient symptoms as shown in Figure 6.

Algorithm	**Accuracy Score Range**	**Notes**
Naive Bayes	80% – 87%	Works well for text/symptom data; assumes feature independence.

Decision Tree	75% – 85%	Easy to interpret; prone to overfitting.
Random Forest	85% – 92%	Ensemble method; handles imbalanced data better.
Support Vector Machine (SVM)	82% – 90%	Effective in high-dimensional spaces; good for classification tasks.
Logistic Regression	70% – 80%	Performs decently with binary/multiclass outputs, but limited with complex patterns.
Artificial Neural Networks (ANN)	88% – 95%	High accuracy with sufficient data; suitable for deep learning approaches.
Ensemble Models	90%+	Combine strengths of multiple models for improved accuracy and robustness.

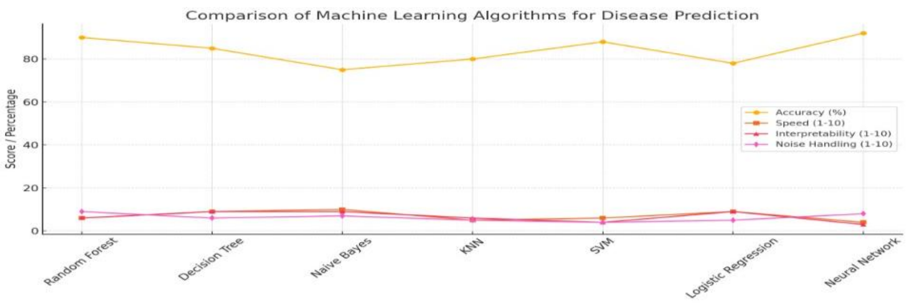

Fig 7: Comparison of Machine Learning Algorithms

Here is a comparison graph showing how various machine learning algorithms perform in terms of accuracy, speed, interpretability, and noise handling for disease prediction systems based on symptoms. Let me know if you'd like a bar chart or any specific metric visualized separately as shown in Figure 7.

Conclusion

The Medicine Recommendation System utilizes machine learning to offer accurate drug suggestions based on user symptoms. By analyzing patient feedback, it enhances healthcare efficiency, reduces prescription errors, and delivers personalized treatment plans. User-friendly features such as a responsive interface and speech recognition boost accessibility.

The system aids both patients and clinicians in making informed decisions, with potential future enhancements including real-time updates and deeper AI insights.

This AI-driven medicine recommendation platform accurately suggests medications based on user symptoms. It boosts efficiency in healthcare, reduces human error, and tailors treatments using patient feedback. Designed with accessibility in mind, the platform includes speech input and an engaging interface. Future developments may include real-time updates and advanced analytics to further empower informed healthcare choices.

This AI-driven medicine recommendation platform accurately suggests medications based on user symptoms. It boosts efficiency in healthcare, reduces human error, and tailors treatments using patient feedback. Designed with accessibility in mind, the platform includes speech input and an engaging interface. Future developments may include real-time updates and advanced analytics to further empower informed healthcare choices.

References

Mayur Gadekar, 2022, June. Symptoms Based Disease Prediction International Conference for Emerging Technology (INCET) (pp. 1-7). IEEE.

Ferjani, M.F., 2023. Disease Prediction Using Machine Learning. Bournemouth, England: Bournemouth University. International Journal of Scientific Research in Science and Technology

MD. Atikur Rahman, Tania Nipa, 2023 April. Predicting disease from several symptoms using machine learning approach. International Journal of Scientific Research in Science and Technology

Mamata Garanayak, 2023. An Intelligent Disease Prediction and Drug Recommendation Prototype by Using Machine Learning. International Journal of Scientific Research in Science and Technology

Keniya, R., Khakharia, 2022. Disease prediction from various symptoms using machine learning. Available at SSRN 3661426.

Dahiwade, D., Patle, G. and Meshram, E., 2022, March. Designing disease prediction model using machine learning approach. IEEE.

Brahmaji Godi, 2022. E-Healthcare Monitoring System using IoT with Machine Learning Approaches.
Marouane Ferjani, 2023. Disease Prediction Using Machine Learning. BMC medical informatics and decision making, International Journal of Scientific Research in Science and Technology 19(1), pp.1-16.

Jai Singh, 2021. Disease Identifier by Symptoms Using Machine Learning. IEEE.

K. Gaurav, 2023. Human Disease Prediction using Machine Learning Techniques and Real-life Parameters.

Kohli, P.S. and Arora, S., 2023, December. Application of machine learning in disease prediction. In 2018 4th International conference on computing communication and automation (ICCCA) (pp. 1-4). IEEE.

Kajal Patil, 2022, November. Multiple Disease Prognostication Based On Symptoms Using Machine Learning Techniques.

Jadhav, S., Kasar, R., Lade, N., Patil, M. and Kolte, S., 2023. Disease prediction by machine learning from healthcare communities. International Journal of Scientific Research in Science and Technology, pp.29-35.

P. Yaswanthi, Vinusha, H. and Sajini, S., 2023. Machine Learning Based Disease Prediction from User Symptoms.(CSEIJ), 8(1), pp.1-8.

Vijiyarani, S. and Sudha, S., 2023. Disease prediction in data mining technique–a survey. International Journal of Computer Applications & Information Technology, 2(1), pp.1721.

https://www.kaggle.com/datasets/kaushil268/disease-prediction-using-machine-learning
https://iq.opengenus.org/gaussian-naive-bayes/
https://www.ibm.com/topics/random-forest