DISEASE PREDICTION THROUGH SYMPTOM ANALYSIS AND IMAGE INTERPRETATION

Authors:

1. **Tabassum Mallikar ,(*BE) Computer Science And Engineering***
Don Bosco Institute Of Technology Bengaluru,India tabassummallikar@gmail.com
2. **Varsha S,(*BE) Computer Science And Engineering***
Don Bosco Institute Of Technology Bengaluru,India varshagowda11082004@gmail.com

Corresponding Author (Guide)
Dr. Sameena Bano Associate professor
Dept of Computer Science And Engineering Don Bosco Institute Of Technology Bengaluru,
sameenabano@dbit.co.in.

Abstract

The rapid advances in artificial intelligence (AI) and machine learning (ML) technologies, particularly in the area of disease prediction, are rapidly altering medical care. The development of a prediction model that employs machine learning and artificial intelligence techniques to analyze patient data and predict the onset of various illnesses is the focus of this project. In an attempt to identify patterns and trends that might best predict the occurrence of diseases like diabetes, heart disease, or liver disease, it is modelled using data that includes demographics, medical history, and symptoms. Models' accuracy is evaluated and compared using techniques such as decision trees, support vector machines,random forests, and neural networks. The goal of this system is to empower medical professionals to make better decisions by utilizing data- based knowledge. In an effort to enhance patient outcomes, the goal is to develop an intelligent system that supports medical professionals in early diagnosis and treatment planning.

Keywords: early diagnosis, machine learning, artificial intelligence, disease prediction, healthcare analytics, as well as patient health information.

INTRODUCTION

Artificial intelligence (AI) and machine learning (ML) have completely transformed the healthcare sector by providing new methods for enhancing the accuracy and effectiveness of disease diagnosis and prognosis.
Strong algorithms and massive databases enable AI and ML to find hidden patterns in medical data that would otherwise go unnoticed.These technologies consider a variety of factors, such as demographic information, medical history, and patient symptoms, to assess an individual's risk of developing a specific disease.The objective is to create an intelligent system that aids healthcare providers in early diagnosis and treatment planning in an attempt to improve patient outcomes.The objective is to develop a prediction system based on a machine learning model that can examine patient data and spot potential health risks.
Methods such as random forests, decision trees, and support vector machines are employed to support potential

health hazards with a high degree of predictive accuracy.

To achieve high prediction accuracy so that medical practitioners can make the best decisions, methods like support vector machines, decision trees, and random forests are employed.

Lastly, this approach seeks to improve patient outcomes and health care management by maximizing early detection, lowering diagnostic errors, and facilitating early medical intervention

LITERATURE SURVEY

Swathi M et al. (2024) created a disease prediction system based on both symptom identification and image processing to aid in diagnosis. The system has a user-friendly interface and supports multiple input types, although it can be prone to limited dataset scope and lesser model accuracy

Dhruv Sharma et al. (2024) introduced a system based on GANs and AI/ML for accurate medical image analysis, disease prediction, and report summarization. The method is highly accurate and improves report generation but requires intricate training and heavy computational requirements.

Hema et al. (2022) created a disease prediction system that employs machine learning algorithms like Naive Bayes, Decision Trees, and Random Forests based on symptoms input by the users. The system has a GUI and tries out several algorithms but is restricted to symptom inputs alone and can be challenged by ambiguous cases or complicated ones.

TABLE 1.1

Author & Year	**Title**	**Purpose**	**Advantages**	**Limitations**
Swathi M et al (2024)	The Disease Prediction System Based on Symptom Identification and Image Processing	To develop a system that predicts diseases using user symptoms and image inputs	Combines image processing and symptoms; user-friendly interface	May lack advanced model accuracy; limited dataset scope
Dhruv Sharma et al (2024)	Medical Image Processing, Disease Prediction and Report Summarization using GANs and AIML	To use GANs and AI/ML for accurate medical image analysis, disease prediction, and automated report generation	High accuracy with GANs; automation of report writing	GANs can be computationally expensive; training complexity

P. Hema et al (2022)	Disease Prediction using Symptoms based on Machine Learning Algorithms	To predict diseases using symptom input and ML algorithms (Naive Bayes, Decision Trees, Random Forests)	Multiple ML algorithms tested; GUI interface	Limited to symptom-based input; may not handle complex cases

COMPARATIVE ANALYSIS

TABLE 1.2

Author(s)	Title	Algorithm(s)	Database	Accuracy
1.Shaurya Gupta, Karan Pal, Deepjyoti Choudhury	An Experimental Analysis of Multiple Disease Prediction using Machine Learning Algorithms	Random Forest, Logistic Regression, Decision Tree, KNN, SVC, AdaBoost, MLP	Diabetes, Kidney Disease, Liver Disease, Asthma	95.8% (Asthma)
2.Swathi M, Sri Aishwariya N, Prithika N G, Jayaram B	The Disease Prediction System Based on Symptom Identification and Image Processing.	Random Forest, Decision Tree, XGBoost, Logistic Regression, LSTM, Naïve Bayes, CNN (DenseNet121)	Multiple datasets	96.5%
3.Sneha Grampurohit, Chetan Sagarnal	Disease Prediction using Machine Learning Algorithms.	Decision Tree, Random Forest, Naïve Bayes	4920 patient records, 41 diseases, 95 symptoms	93%
4.Chaimaa Boukhatem et al.	Heart Disease Prediction Using Machine Learning	MLP, SVM, RF, NB	Patient health records	91.6%

5.Garima Shukla et al.	Chronic Kidney Disease Prediction Using Machine Learning Algorithms and the Important Attributes for the. Detection.	KNN, Decision Tree, ANN	CKD dataset	98.6%
6.Uday Kiran Kommineni et al.	Human Disease Prediction based on Symptoms	Random Forest, Naïve Bayes	Preprocessed symptom dataset	89.4%
7.Hosam F. El-Sofany	Predicting Heart Diseases Using Machine Learning and Various Data Classification Methods	NB, SVM, Voting, XGBoost, AdaBoost, Bagging, DT, KNN, RF, LR	Public and private heart disease datasets	97.5%
8.G. Parimala et al.	Diabetes Prediction using Machine Learning	KNN, Label Encoder, Train- test Split (ensemble techniques)	Diabetes dataset	96%
9.Kalpesh Vinayak Joshi et al.	Sym-Diagnose: Symptoms based Disease Diagnosis & Healthcare Suggestion	ML models + NLP (Gemini API)	Symptom-based dataset	78%
10.Nidhi Kosarkar et al.	Disease Prediction using Machine Learning	Decision Tree, Random Forest, Naïve Bayes	Kaggle dataset	98%

FIGURES

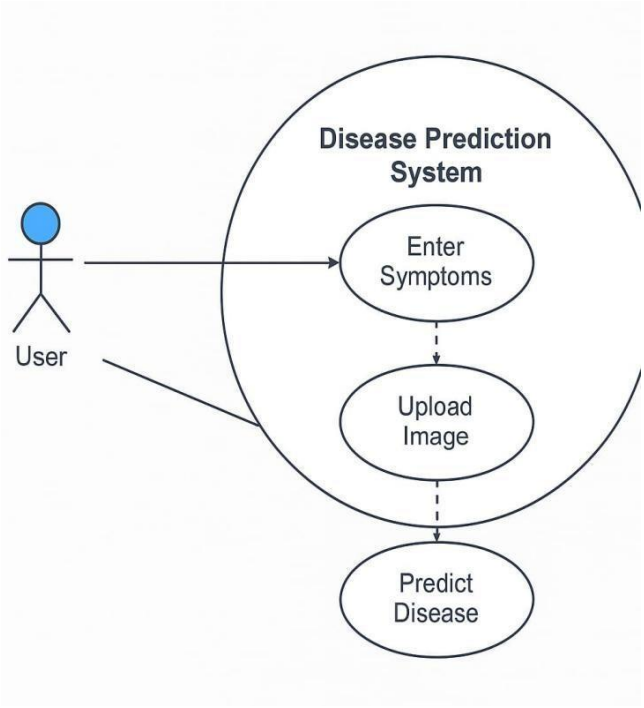


Figure 1. Use case diagram

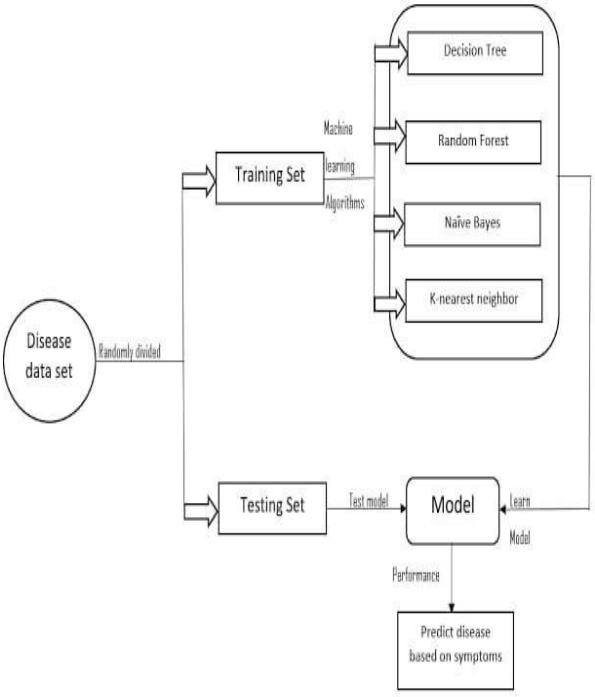


Figure 2. Flow diagram

RESULTS/DISCUSSION

Different machine learning models were tried for disease prediction, and they are Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and k-Nearest Neighbors (KNN). Of these, models such as Random Forest and Decision Tree were found to be highly accurate and could be applied for early classification of health data based on symptoms, demographics, and history. The data was preprocessed and divided in an 80:20 manner for training and testing, so that balanced evaluation is carried out. Feature selection applied correlation analysis and rank importance to enhance efficiency and concentrate on major indicators. Performance was assessed with accuracy, precision, recall, F1-score, confusion matrix, and ROC-AUC, enabling overall comparison. The ultimate model demonstrated high predictive capability and was applied for real-time or batch predictions. It was successful in detecting conditions such as diabetes, heart conditions, and liver conditions, enabling early action. It is also used to access remotely, so users can receive predictions without going to clinics. The challenges are ensuring privacy and enhancing comprehension of complex models. Overall, the research indicates that machine learning assists in enhancing disease prediction and informing timely decisions.

CONCLUSION

The Disease Prediction Using AIML system uses machine learning to effectively predict diseases from patient health information, including symptoms, medical history, and lifestyle. Through early detection, the system facilitates timely intervention, which lessens the severity of the diseases. Its automation gives it consistent and accurate predictions, while scalability enables ongoing improvement with new data being integrated.The system also increases accessibility, where patients can enter data remotely and obtain predictions without the need for urgent doctor consultations. While issues such as data privacy and interpretability of models persist, the accuracy of the system will increase with continued improvement in machine learning.This remedy is a major milestone on the path to predictive healthcare, promising the possibility of customized, effective care and better patient outcomes. With additional tuning and more integration of data, its influence on the health care industry will expand.

REFERENCES

[1] Swathi M, Sri Aishwariya N, Prithika N G, and Jayaram B, “The Disease Prediction System Based on Symptom Identification and Image Processing,” 2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE), pp. 1075–1079, Apr. 2024.

[2] Dhruv Sharma, Abdus Subhan, Shuchi Singh, and Meenakshi R. Patil, “Medical Image Processing, Disease Prediction and Report Summarization Using Generative Adversarial Networks and AIML,” 2024 2nd World Conference on Communication & Computing (WCONF), Raipur, India, Jul. 12–14, 2024.

[3] Shaurya Gupta, Karan Pal, and Deepjyoti Choudhury, "An Experimental Analysis of Multiple Disease Prediction Using Conference on Computational Intelligence and Communication Networks (CICN), pp. 200–206, 2024.

[4] Kalpesh Vinayak Joshi, Rohit Kishor Jadhav, Hardik Vasant Rokde, Shrey Santosh Rupnavar, Ruchi Pankaj Choudhary, Rohit Sharad Kalaskar, and Bhargav Milind Rokade, "Sym-Diagnose: Symptoms based disease Diagnosis & Healthcare Suggestion," 2024 International Conference on Artificial Intelligence and Quantum Computation-Based Sensor Application (ICAIQSA), pp. 1–8, 2024.

[5] Sneha Grampurohit and Chetan Sagarnal, "Disease Prediction using Machine Learning Algorithms," 2020 International Conference for Emerging Technology (INCET), Belgaum, India, Jun. 5–7, 2020, pp. 1–7.

[6] Nidhi Kosarkar, Pallavi Basuri, Poonam Karamore, Prachi Gawali, Pradnya Badole, and Pranjali Jumle, "Disease Prediction using Machine Learning," 2022 10th IEEE International Conference on Emerging Trends in Engineering & Technology - Signal and Information Processing (ICETET-SIP-22), 2022.

[7] Chaimaa Boukhatem, Heba Yahia Youssef, and Ali Bou Nassif, "Heart Disease Prediction Using Machine Learning," 2022 Advances in Science and Engineering Technology International Conferences (ASET), 2022.

[8] G. Parimala, R. Kayalvizhi, and S. Nithiya, "Diabetes Prediction using Machine Learning," 2023 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, Jan. 2023.

[9] Garima Shukla, Gaurav Dhuriya, Sofia K. Pillai, and Aradhna Saini, "Chronic Kidney Disease Prediction UsinG Machine Learning Algorithms and the Important Attributes for the Detection," 2023.

[10] Uday Kiran Kommineni, Dhatri Phani Priya Kowthavarapu, Kalyan Chakravarti Yelavarti, and Guna Sri Manjunadh Polukonda, "Human Disease Prediction based on Symptoms," 2023 7th International Conference on Computing Methodologies and Communication (ICCMC), 2023.

[11] Farha Deeba and Siddarama R. Patil, "Implementation of Artificial Intelligence in Disease Prediction and Healthcare System – A Survey," 2021 Innovations in Power and Advanced Computing Technologies (i-PACT), 2021.

[12] Yogesh Kumar, Apeksha Koul, Ruchi Singla, and Muhammad Fazal Ijaz, "Artificial Intelligence in Disease Diagnosis: A Systematic Literature Review, Synthesizing Framework and Future Research Agenda," Journal of Ambient Intelligence and Humanized Computing, vol. 14, pp. 8459–8486, 2023.

[13] Swathi M, Sri Aishwariya N, Prithika N G, and Jayaram B, "The Disease Prediction System Based on Symptom Identification and Image Processing," 2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE), pp. 1075–1078, 2024, doi: 10.1109/IC3SE62002.2024.10593340.

[14] Dhruv Sharma, Abdus Subhan, Shuchi Singh, and Meenakshi R. Patil, "Medical Image Processing, Disease Prediction and Report Summarization using Generative Adversarial Networks and AIML," 2024 2nd World Conference on Communication & Computing (WCONF), Raipur, India, pp. 1–3, Jul. 2024, doi: 10.1109/WCONF61366.2024.10692015

[15] M. Ramya, Meenakshi B K, and Kanagalakshmi M, "Optimizing Symptom and Disease Representation with Structural and Textual Information Integration," Proceedings of the International Conference on Multi-Agent Systems for Collaborative Intelligence (ICMSCI- 2025), pp. 1264–1267, 2025, doi: 10.1109/ICMSCI62561.2025.10893971.

Copyright & License:
© Authors retain the copyright of this article. This work is published under the Creative Commons Attribution 4.0 International License (CC BY 4.0), permitting unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.