LITERATURE SURVEY

REAL-TIME DISEASE PREDICTION WITH SYMPTOMS USING MACHINE LEARNING

Submitted to :

Ms. Arshi Riyaz
Assistant Professor
Dept. of Technology

Submitted by :

Laxit Jangid
22EJIAI053
Parthivi Khatri
22EJIAI072

Department of Computer Science Engineering (AIML)
Jodhpur Institute of Engineering and Technology
Session 2025-26

ABSTRACT

Healthcare is one of the most essential needs, yet access to timely medical advice is still limited for many people. Often, individuals experience symptoms but hesitate to visit a doctor immediately, either due to distance, cost, or lack of awareness. This project focuses on building a simple yet effective disease prediction system that can give users an early indication of possible illnesses based on the symptoms they provide. The approach uses text-processing methods like TF-IDF to convert symptoms into numerical form and applies a Random Forest classifier to generate predictions. Unlike heavy deep learning models, our system is intended to stay lightweight so that it can run smoothly on common devices in real time. We also plan to include multilingual support so that people from different backgrounds can use it without difficulty. The aim is not to replace doctors but to provide a supportive tool that encourages users to seek proper medical consultation at the right time.

INTRODUCTION

This project aims to develop a real-time disease prediction system that uses machine learning algorithms to analyze user-input symptoms and suggest possible diseases. The system will leverage a pre-trained dataset of symptom-disease relationships to provide instant predictions. Users can enter their symptoms through a web or mobile interface, and the model will generate potential diagnoses with confidence levels. This tool is designed to serve as a preliminary health-check system, providing early guidance and encouraging timely medical consultation, thereby reducing the burden on healthcare professionals and improving accessibility, particularly in remote areas. The project will involve data collection and preprocessing, model selection (such as Decision Tree, Random Forest, Naïve Bayes, or Deep Learning), training, real-time processing, and performance evaluation using metrics like accuracy, precision, recall, and F1-score. The final system will be deployed on a cloud platform like AWS or Firebase for real-time access.

RELATED WORK

1. **GeeksforGeeks. " Disease Prediction Using Machine Learning."**
   https://www.geeksforgeeks.org/machine-learning/disease-prediction-using-machine-learning/

This website provided a practical, step-by-step guide for implementing a disease prediction system using machine learning. It was particularly useful for understanding the end-to-end process, including data preprocessing, model training, and evaluation using various classifiers like SVC, Naïve Bayes, and Random Forest.

2. **Jmarihawkins. "Disease_Prediction_App: The Disease Prediction Project."**
   https://github.com/jmarihawkins/Disease_Prediction_App

This GitHub repository was instrumental in understanding the technology stack and methodology for a similar project. It detailed the use of libraries like Pandas, Scikit-learn, and TensorFlow/Keras, and provided a comparative analysis of different models (MLP, CNN, LSTM) in the context of disease prediction based on symptoms.

3. **Vatshayan. "Final-Year-Disease-Prediction-Project."**
   https://github.com/Vatshayan/Final-Year-Disease-Prediction-Project

This repository provided insights into a final-year project on disease prediction. It was helpful in understanding the system architecture and included a link to an accompanying research paper, which served as a foundational resource.

4. **Disease Prediction using Machine Learning, Deep Learning and Data Analytics**

This book provides a comprehensive review of healthcare technologies and data analytics, with Chapter 3 specifically focusing on machine learning (ML) and deep learning (DL) techniques for detecting disease severity. It was particularly helpful for the project as it offered insights into methodologies and algorithms for analyzing medical data, enabling the prediction of diseases and assessment of disease severity for personalized treatment.

5. M*Authors: G. Rani, V. S. Dhaka, and P. K. Tiwari*achine Learning for Disease Detection, Prediction, and Diagnosis: Challenges and Opportunities

Authors: *T. Choudhury and A. Katal*

This book explains modern techniques for improving ML practices in healthcare, covering supervised learning, deep neural networks, and convolutional neural networks (CNNs). It served as a valuable reference for developing methodologies for disease prediction using patient health data. Additionally, it helped in understanding supervised learning methods and deep learning models, which were important for disease detection and prediction in this project.

6. Practical Data Mining Techniques and Applications (Chapter: Application of Machine Learning in Disease Prediction)

Authors: *S. Rao, S. Kulkarni, S. Mehta, and N. Katre*

This book discusses the application of machine learning for predicting chronic diseases such as diabetes, heart disease, and breast cancer. It emphasizes the use of the Naïve Bayes (NB) classifier for accurate predictions. For this project, it provided useful methodological guidance for building disease prediction models using patient health data and NB classifiers, making it a strong reference for practical implementation.

7. **Diagnosis of Liver Disease using Machine Learning Models**

Publisher/Conference: *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)*

This paper compares machine learning models such as SVM, Decision Tree, and Random Forest for predicting liver disease. It provides a detailed discussion of precision, accuracy, and recall as key evaluation metrics. For this project, the paper guided the evaluation phase by offering useful insights into performance metrics and algorithm comparisons for disease prediction.

8. Disease Prediction by Machine Learning over Big Data from Healthcare Communities

Publisher/Conference: *IEEE Access, vol. 5, 2017*

This research explores the application of machine learning for disease prediction using large-scale healthcare datasets. It emphasizes big data handling and real-time analysis of patient health information. In this project, the paper helped design the

system to effectively manage large datasets and provided theoretical foundations for accurate disease prediction.

9. Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques

***Publisher/Conference**: IEEE Access, vol. 7, 2019*

This paper proposes a hybrid machine learning model that combines multiple algorithms to improve the accuracy of heart disease prediction. It highlights how hybrid approaches enhance performance compared to using individual classifiers. The study inspired this project to integrate hybrid techniques, thereby improving accuracy and overall performance in disease prediction models.

CONCLUSION

From the studies and resources reviewed, it is clear that there is still a big gap in having a disease prediction system that is both real-time and lightweight while also supporting multiple languages. Many existing works are either too simple, like those using only Naïve Bayes or Decision Trees, or too heavy, depending on deep learning models that need large datasets and high computing power. In this project, we tried to fill that gap by using TF-IDF for feature extraction and a Random Forest Classifier, which gives a good balance between accuracy and efficiency. We also focused on making the system multilingual, starting with English and Hindi, so that it is easier for people from different backgrounds to use. Since the system is deployed for real-time use, it can be accessed quickly and practically in daily life.

At the same time, we see many ways to improve this work in the future. The dataset can be expanded to cover more diseases, and deep learning methods can be added to push the accuracy even higher. Another important direction is building a proper mobile or web application so that the system can reach more people directly. Finally, connecting it with IoT devices or wearable health sensors would make it even more powerful, as it could provide continuous monitoring instead of only symptom-based input. Overall, this project takes a step toward accessible healthcare support and opens up a path for many useful improvements in the future.

LIST OF REFERENCES

[1] GeeksforGeeks, “Disease Prediction Using Machine Learning.” [Online]. Available: https://www.geeksforgeeks.org/machine-learning/disease-prediction-using-machine-learning/. [Accessed: 16-Sep-2025].

[2] J. M. Hawkins, “Disease_Prediction_App: The Disease Prediction Project.” GitHub Repository. [Online]. Available: https://github.com/jmarihawkins/Disease_Prediction_App. [Accessed: 16-Sep-2025].

[3] Vatshayan, “Final-Year-Disease-Prediction-Project.” GitHub Repository. [Online]. Available: https://github.com/Vatshayan/Final-Year-Disease-Prediction-Project. [Accessed: 16-Sep-2025].

[4] G. Rani, V. S. Dhaka, and P. K. Tiwari, *Disease Prediction using Machine Learning, Deep Learning and Data Analytics*. Springer, 2021.

[5] T. Choudhury and A. Katal, *Machine Learning for Disease Detection, Prediction, and Diagnosis: Challenges and Opportunities*. CRC Press, 2020.

[6] S. Rao, S. Kulkarni, S. Mehta, and N. Katre, *Practical Data Mining Techniques and Applications*, Chapter: Application of Machine Learning in Disease Prediction. Springer, 2019.

[7] R. Shyamala and M. A. Kumar, “Diagnosis of Liver Disease using Machine Learning Models,” in *Proc. 2020 Fourth Int. Conf. on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)*, Palladam, India, pp. 123–128, 2020.

[8] R. Kumar and S. Reddy, “Disease Prediction by Machine Learning over Big Data from Healthcare Communities,” *IEEE Access*, vol. 5, pp. 8869–8879, 2017.

TEAM :-

LAXIT JANGID
PARTHVI KHATRI