Disease Prediction Based On Symptoms

Harry Raj
210962022
harry.raj@learner.manipal.edu
Roll no: 9

Harika Boppana
210962030
harika.boppana@learner.manipal.edu
Roll no: 13

I. INTRODUCTION:

The integration of Big Data Analytics in healthcare has revolutionized disease prediction. This project focuses on leveraging BDA techniques to predict diseases based on reported symptoms. By analyzing extensive datasets, including symptomatology, the aim is to construct predictive models for early detection and intervention, improving patient outcomes and enabling proactive healthcare measures.The integration of Big Data Analytics into healthcare has transformed disease prediction, yet mental health disorders remain challenging due to the abstract nature of symptoms and the need for comprehensive analysis.

This project addresses this gap by leveraging BDA techniques to predict mental health disorders based on patient symptoms. Despite the complexities involved, accurately recording and analyzing symptoms can lead to informed conclusions about a patient's mental health.

Our approach utilizes Decision Trees, Naive Bayes, and Random Forest classifiers, supported by PySpark for efficient data analysis and prediction. Recognizing the pivotal role of data in mental health, our project aims to provide valuable insights into patients' mental states, offering opportunities for improved accuracy and manageability in mental healthcare.

By harnessing big data technologies, we aim to unlock the full potential of data-driven approaches, paving the way for enhanced mental health diagnosis and treatment.

II. LITERATURE REVIEW

The paper "Disease Prediction using Machine Learning Algorithms" presented at the 2020 International Conference for Emerging Technology discusses the application of machine learning techniques in healthcare for disease prediction. Using a dataset of 4920 patient records with 132 symptoms associated with 41 diseases, the study aims to develop a prediction model using machine learning algorithms such as Decision Tree, Random Forest, and Naive Bayes. Data preprocessing techniques were applied to

select 95 relevant symptoms to avoid overfitting. The study demonstrates the implementation and comparative analysis of these algorithms, showing promising results with an accuracy of up to 95%. However, limitations such as the risk of overfitting when considering all 132 symptoms and the need for further validation on larger datasets should be addressed in future research. Additionally, while the paper provides insights into the potential of machine learning in disease prediction, practical implementation and usability of the developed system in clinical settings remain areas for exploration in the literature.

the paper presents a Disease Prediction system utilizing Machine Learning algorithms to forecast diseases based on user-provided symptoms. Leveraging the Naïve Bayes classifier and other algorithms like Decision Tree and Logistic Regression, the system achieves high prediction accuracy for diseases such as Diabetes, Malaria, Jaundice, Dengue, and Tuberculosis. Performance evaluation metrics such as recall, precision, accuracy, and F1 measures are utilized to assess the system's effectiveness. However, while the system demonstrates promising results, several limitations should be considered in future research. Firstly, the system's performance may vary depending on the diversity and quality of the input data, highlighting the importance of comprehensive and representative datasets. Secondly, the reliance on user-provided symptoms may introduce biases or inaccuracies, particularly if users omit relevant information or provide misleading data. Additionally, the system's effectiveness in real-world clinical settings and its ability to handle complex and rare diseases warrant further investigation. Finally, scalability and computational efficiency may pose challenges when deploying the system in large-scale healthcare environments, necessitating optimization strategies for improved performance. These limitations underscore the need for ongoing research to enhance the robustness and applicability of disease prediction systems in clinical practice.

The paper introduces a Disease Prediction system using Machine Learning algorithms, aiming to forecast diseases based on user-provided symptoms. Leveraging algorithms such as Random Forest, Naive Bayes, and Support Vector Machine, the system provides accurate predictions, offering users insights into their health conditions without the need for immediate medical consultation. However, several drawbacks need consideration. Firstly, the system's effectiveness heavily relies on the quality and diversity of the input data, which may vary significantly and affect prediction accuracy. Secondly, the challenge of understanding medical terminology among laymen users poses a barrier to accurate symptom reporting, potentially leading to misdiagnoses. Additionally, while the system boasts high accuracy rates, the practical applicability and scalability in real-world healthcare settings require further exploration. Integrating the system seamlessly into clinical workflows and ensuring its reliability across different

patient demographics and medical conditions are essential for its widespread adoption and effectiveness. Moreover, future research should focus on addressing these limitations and refining the system's predictive capabilities to enhance its utility and reliability in healthcare decision-making processes.

This paper introduces a disease prediction system based on symptoms, leveraging machine learning, particularly the Random Forest algorithm, to aid in early disease identification. Emphasizing the importance of timely healthcare interventions amidst busy lifestyles, the system allows users to input symptoms via a Python-based graphical user interface (GUI) developed using Tkinter. Training on a dataset containing symptoms and diseases, the model achieves a 95.12% accuracy during testing. However, the paper lacks a thorough evaluation of real-world applicability and future work, only briefly mentioning potential enhancements to user interface and accuracy without detailed strategies for implementation. Further research addressing scalability, robustness, and usability in clinical settings would provide valuable insights for system development and deployment.

III. RESEARCH GAPS

1. While BDA is widely used in healthcare for disease prediction, its application to mental health disorders remains underexplored due to the abstract nature of symptoms. Research in this area could provide insights into early intervention and treatment for mental health conditions.

2. Existing research lacks comprehensive validation and usability studies in real-world clinical settings, hindering the scalability and reliability of disease prediction systems. Future research should focus on extensive validation to assess system usability and reliability in clinical practice.

3. Disease prediction systems relying on user-provided symptoms may encounter biases or inaccuracies. Future research should address this challenge by exploring techniques to improve data accuracy and reliability, especially when users provide incomplete or misleading information.

IV. OBJECTIVES

1. Analyse patient symptoms thoroughly and create an insightful medical record from their history. This helps doctors get a better understanding of the patient's present health condition based on their past conditions/symptoms/behaviours.

2. Make an accurate prediction based on synptoms and other lifestyle factors of patient. This can aid medical professionals in coming to a conclusive diagnosis.

3. Help doctors draw a treatment plan and preventative care based on how the patient's data changes with different medication and treatment methods. Doctors can use data to see how patient reacts to certain treatment plans as it reflects in their data.

4. Build an interactive and easy to use user interface where patients/doctors can use the application by entering their symptoms and receiving analysis and predictions.

V. REFERENCES

[1]S. Grampurohit and C. Sagarnal, "Disease Prediction using Machine Learning Algorithms," 2020 International Conference for Emerging Technology (INCET), Belgaum, India, 2020, pp. 1-7, doi: 10.1109/INCET49848.2020.9154130. keywords: {Diseases;Decision trees;Entropy;Forestry;Classification algorithms;Pain;Data mining;Machine Learning;Data mining;Decision Tree classifier;Random forest classifier;Naive Bayes classifier},

[2]Gomathy, C K. (2021). THE PREDICTION OF DISEASE USING MACHINE LEARNING.

[3]https://www.irjmets.com/uploadedfiles/paper/issue_5_may_2022/24065/final/fin_irjmets1653367944.pdf

[4]K. S. Kumar, M. Sai Sathya, A. Nadeem and S. Rajesh, "Diseases Prediction based on Symptoms using Database and GUI," 2022 6th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2022, pp. 1353-1357, doi: 10.1109/ICCMC53470.2022.9753707. keywords: {Databases;Medical services;Forestry;Big Data;Random forests;Diseases;Graphical user interfaces;Big data analytics;Disease prediction and diagnosis;Random Forest;Database;Graphical User Interface (GUI)},