International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 08 Aug. 2024, pp: 34-39 www.ijaem.net ISSN: 2395-5252

Predicting Diseases and Prescribing Medicines: A Comparative Analysis of Machine Learning Techniques

Chandana H R[*1], Ms. Ashwini C[*2]
P.G, Student, Department of Master of Computer Applications, University B.D.T Collage of Engineering, Davangere, Karnataka, India
Assistant Professor, Department of Master of Computer Applications, University B.D.T Collage of Engineering, Davangere, Karnataka India

Date of Submission: 01-08-2024 Date of Acceptance: 08-08-2024

ABSTRACT: Many people still rely on traditional medical care methods, like visiting hospitals and seeing doctors for even minor symptoms, even in the era of easily accessible advanced technology and easy access to information. However, this approach can be time-consuming and inefficient, so research based on alternative methods is proposed. The objective is to create a prediction model that, using patient symptoms as input, can reliably identify probable diseases. By analyzing and interpreting symptom data using machine learning, this approach makes disease prediction quicker and more precise. The research demonstrates a high degree of accuracy in disease prediction, and the model is trained and validated using relevant patient features and medical record data. The study's conclusions indicate that the suggested model may prove to be helpful in the early detection and management of diseases, therefore improving patient outcomes. In summary, this research signifies a significant breakthrough in the field of health informatics, with potential advantages for the prevention, diagnosis, and treatment of diseases.

Keywords: Machine Learning, Dataset, KNN, SVM, GUI.

I. INTRODUCTION

Global disease incidence is rising, which puts a heavy burden on healthcare organization and emphasizes the importance of early and precise diagnosis for efficient management and treatment. Conventional diagnostic techniques, which mainly rely on a great deal of medical training and experience, can be resource and time-intensive, frequently resulting in delays that worsen patient conditions and raise healthcare expenses. Machine learning (ML) presents itself as a viable remedy in this situation, with the ability to speed up and automate the diagnostic procedure. Medical practitioners' diagnostic abilities are enhanced and patients are given early assessments, which can analyze massive datasets, spot complex patterns, and make predictions based on input features like symptoms. In this research, a complete disease prediction system based on user-reported symptoms is presented. It makes use of multiple ML techniques. Using a graphical user interface (GUI) that is simple to use and was created with Tkinter, users may input their symptoms and receive a predicted diagnosis along with comprehensive details including descriptions, precautions, drugs, workouts, and diet plans. This study aims to create and evaluate various ML algorithms for disease prediction. Our system offers you to select the model and it also allow you to input the symptoms and it will predict the disease based on the symptoms. In order to create a robust dataset, this project will collect data from a selection of sources. It will also preprocess and select features, train and evaluate multiple machine learning models to identify the diseases, and design and implement a graphical user interface (GUI) for user interaction.

II. LITERATURE REVIEW

As a result of multiple studies showing how ML may improve diagnostic speed and accuracy, the use of this technology in disease prediction has attracted a lot of attention recently. Predicting diseases like diabetes, heart disease, and cancer has been a successful use of decision trees(DTrees), which are renowned for their simplicity and comprehension. Such uses have, for example, been made possible by R. Quinlan's ID3 and C4.5 algorithms. Because of its creativity and ability to handle large amounts of data, support vector machines, or SVMs, have proven highly accurate when used for disease prediction tasks like breast cancer diagnosis. Through the combination of several decision trees, Breiman's ensemble learning technique known as Random Forests(RFs) has demonstrated promise in the classification of medical imaging data and the

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 08 Aug. 2024, pp: 34-39 www.ijaem.net ISSN: 2395-5252

prediction of cardiovascular diseases. In numerous research on disease prediction, including the prediction of the beginning of diabetes, Friedman's Gradient Boosting Machines (GBMs) have shown excellent predictive accuracy. Patterns in patient data have been found for heart disease prediction using the straightforward yet powerful K-Nearest Neighbors (KNN) algorithm. Because it is capable of both continuous and categorical data, the Naive Bayes classifier—which is based on Bayes' theorem—has been used successfully in the prediction of cardiovascular disease. It is easier to access and utilization of ML models when they are integrated with graphical user interfaces (GUIs). Research has produced easy-to-use instruments for predicting diseases, like graphical user interfaces (GUIs) for diabetes and heart disease prediction that let users enter symptoms and get pertinent data. Training and testing these models depend on the availability of large datasets that include symptoms, descriptions, safety measures, prescription drugs, exercises, and diets. By capturing the complex structure of diseases and improving the models' generality and robustness, these datasets enable the production of more accurate forecasts. Using the advantages of several ML algorithms and a well-thought out graphical user interface, this study seeks to build upon these fundamental studies in order to create an all encompassing and easily comprehensible disease prediction system.

III. METHODOLOGY:

Step1: Data Collection and Preprocessing

Our project includes six CSV files which contain the details on a different facet of diseases—symptoms, descriptions, and also precautions to take, prescription drugs, workout routines, and diets—make up the dataset utilized for this study. Before being processed further to guarantee purity and normality, these datasets were assembled from reputable medical sources. Categorical data was encoded where needed, and missing values were dealt with. Binary encoding was utilized to indicate whether symptoms were present in the input vector or not, and each symptom was given a unique identification to aid in processing.

Step2: Model selection and training

The following six machine learning methods were chosen for this investigation:

• Support Vector Machine (SVM) • Decision Tree Classifier

• K-Nearest Neighbors (KNN) • Random Forest Classifier

• Gradient Boosting Classifier

• Multinomial Naive Bayes (NB) It is possible to compare the performance of these models in-depth because of the broad range of categorization strategies that led to their selection. Symptoms were used as input features and diseases as target labels in the training of the models utilising the symptoms dataset.

Step3: Model Evaluation

In order to assess each model's performance, we considered two key factors: their accuracy and their performance on a confusion matrix. In order to prevent overfitting—that is, to prevent the models from becoming too skilled at handling the training data and becoming incapable of handling new information—we also took care to include cross-validation. We trained them on one set of data and then evaluated their performance on an entirely different set to really put them to the test. In this manner, we could assess their capacity for generalization and situational awareness.

Step4: Integration into GUI To enable users to use the disease prediction system, a graphical user interface (GUI) was created using Tkinter. Users can choose a ML model for prediction by entering their symptoms into the GUI.

Key features of GUI Includes the following:

Model Selection: Users have the option to choose from a variety of ML models, like Decision Tree, SVM, Random Forest, Gradient Boosting, KNN, and Multinomial NB.

Symptom Input: Users are able to enter symptoms in a manner separated with commas.

Prediction Display: Detailed information on the disease, including a description, precautions, medication schedules, workout routines, and diet plans, are presented alongside the predicted disease.

Step5: Helper function for detailed information

A dedicated assistance feature was developed to offer comprehensive details regarding the anticipated disease. This feature looks through pertinent databases to find the most important information, such as a description of the disease, preventive measures, suggested drugs, workout routines, and nutrition guidelines appropriate to the disease.

Step6: Model Loading and Prediction

To make loading them easier during runtime, these models were trained and saved as pickle files. The selected model will be loaded by load_model function , whereas the function get_predicted_value makes a disease prediction

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 08 Aug. 2024, pp: 34-39 www.ijaem.net ISSN: 2395-5252

using the input symptoms. Then, using the helper function, detailed information is retrieved based on the expected disease.

Step7: User Interaction and Output
Users are guided through a straightforward process by the system's intuitive UI. Select a model, enter your symptoms, and receive a diagnosis. You are given a comprehensive understanding of the situation and actionable recommendations for the future when the results are provided in an understandable and straightforward manner.

IV. SYTEM DESIGN:

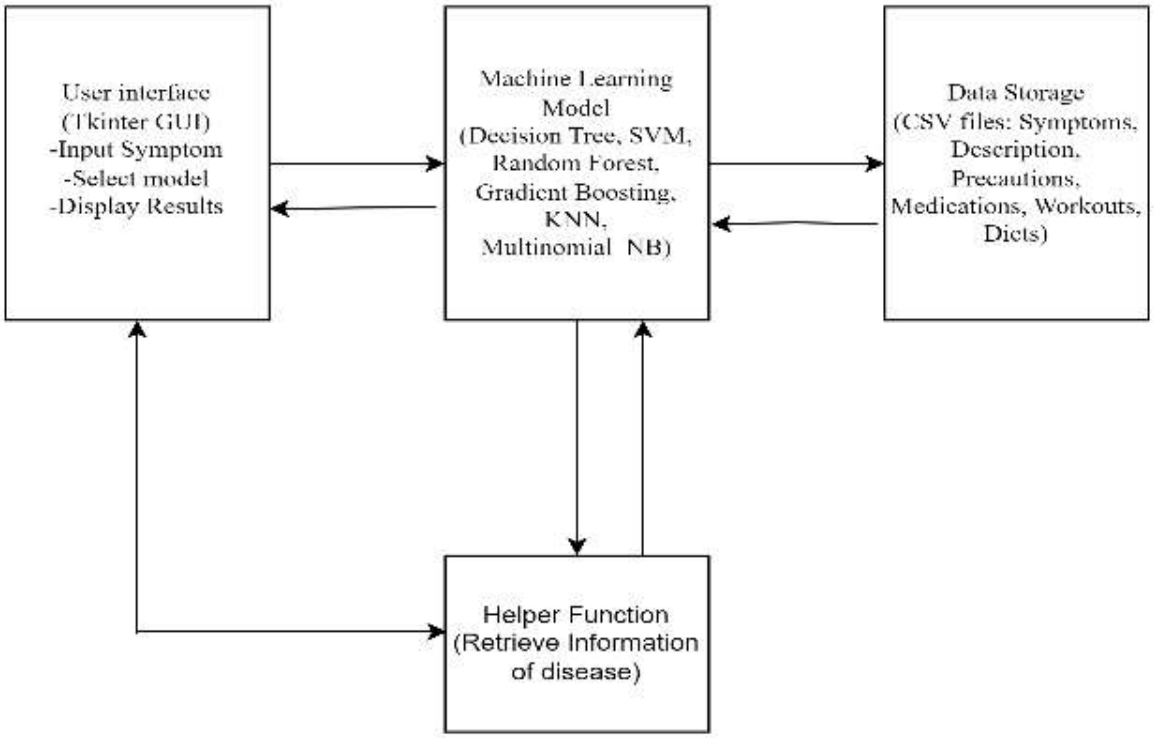

Fig 1.1: System Design

"A user interface, ML models, and data storage are the three primary parts of our disease prediction system.
Users can input symptoms and choose a model using the interface. After then, a full informational graphic and the anticipated disease are displayed.
Symptoms, diseases, and suggested courses of action are all provided by the data storage, which loads pre-trained models based on user decision.
The system is an important tool for managing and predicting diseases because it integrates comprehensive data, precise forecasts, and easy-to-use input."

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 08 Aug. 2024, pp: 34-39 www.ijaem.net ISSN: 2395-5252

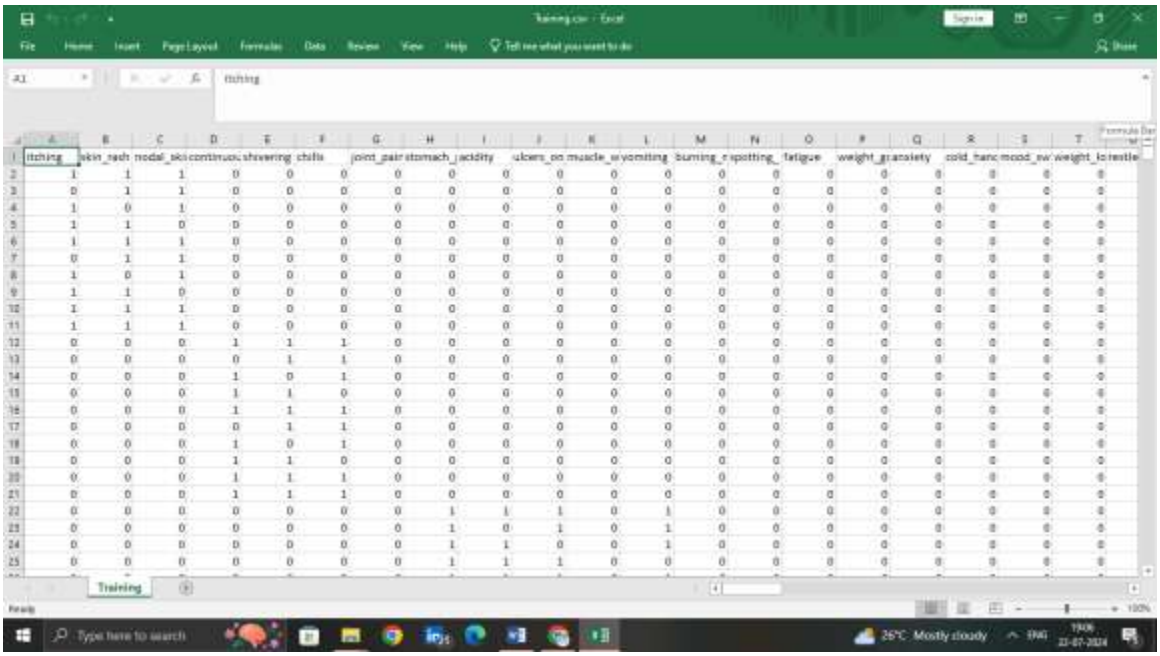

Fig 1.2: Training Dataset

Dataset	Number of Records	Number of Features	Feature Types
Symptoms	4920	132	Binary (0 or1)
Description	41	2	Text
Precautions	41	5	Text
Medications	41	Variable	Text
Workouts	41	Variable	Text
Diets	41	Variable	Text

Fig 1.3: Dataset Overflow

Our disease prediction system makes use of a large dataset of medical records, each containing a distinct set of symptoms identified as present or absent, in order to forecast diseases based on symptoms. This collection includes many different symptoms and their accompanying disorders, each of which is given a numerical code. We've also assembled a number of supporting datasets to give a more thorough picture of each disease, such as thorough disease descriptions, advised warnings, often given medications, helpful workouts and exercises, and customized diets recommendations. We prepare the dataset by completing any missing information, standardizing the data to assure consistency, and translating categorical data into numerical formats before our ML models are able to produce reliable predictions.

V. RESULT AND DISCUSSION

We conducted experiments with our disease prediction technology, and the outcomes

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 08 Aug. 2024, pp: 34-39 www.ijaem.net ISSN: 2395-5252

are remarkable. We used a carefully selected dataset of patient reported symptoms to train and assess a variety of ML models, including Decision Tree, SVM, Random Forest, Gradient Boosting, K-Neighbors, and Multinomial Naive Bayes. We used metrics like accuracy scores and confusion matrices to obtain a comprehensive view of each model's performance.According to the findings, several models outperformed the others and had greater accuracy rates, in particular Random Forest and Gradient Boosting. This indicates that they have exceptional ability to handle intricate symptom patterns and accurately forecast diseases. That's not all, though. We went one step further and used Tkinter to create an intuitive user interface where users can enter their symptoms and receive immediate recommendations. Not only that, but our technology also offers comprehensive details on the disease that is anticipated, such as descriptions, safety measures, prescriptions, workout , and diet suggestions. With this all-encompassing approach, customers are guaranteed not only precise disease prognoses but also practical guidance on how to properly manage their condition.To sum up, the disease prediction technology has shown its value in actual situations, proving to be dependable and practically applicable.

Fig 1.3: Developed GUI

Fig 1.4: Predicted Result I

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 08 Aug. 2024, pp: 34-39 www.ijaem.net ISSN: 2395-5252

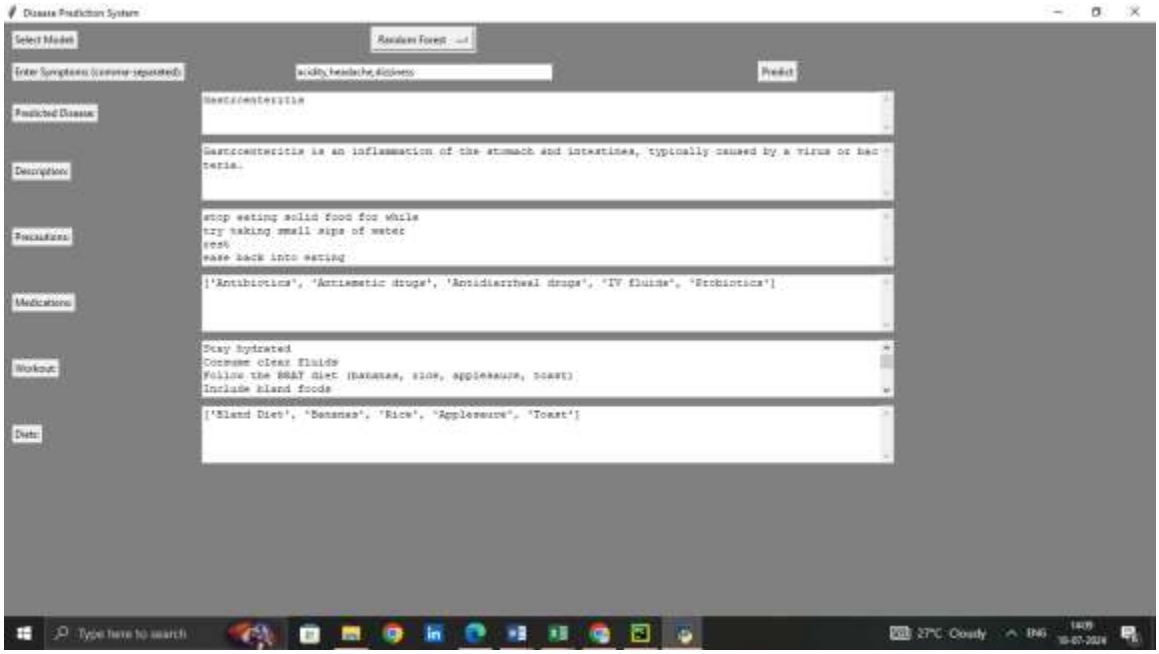

Fig 1.5: Predicted Result II

VI. CONCLUSION

With the use of many classifiers and ML algorithms, our disease prediction system reliably predicts diseases based on patient-reported symptoms, offering users clear and actionable health insights. Providing users with holistic insights such as disease descriptions, precautions, medications, workouts, and diets, our technology goes above and beyond simple forecasts to deliver comprehensive information. The system we have developed is user-friendly and accessible to a wide range of users. This shows how ML may transform the healthcare industry and pave the way for future developments in automated diagnostic systems. We have developed an approach that can enhance early disease identification and management by fusing data science, healthcare, and software development.

REFERENCES

[1]. .https://www.researchgate.net/publication/379662123_Disease_Prediction_and_Medicine_Recommendation_Systems_A_Comparative_Analysis_on_learning_algori thms
[2]. https://www.researchgate.net/publication/373835014_An_Intelligent_Disease_Prediction_and_Drug_Recommendation_Prototype_by_Using_Multiple_Approaches_of_Machine_Learning_Algorithms
[3]. https://iaeme.com/MasterAdmin/Journal_uploads/IJARET/VOLUME_12_ISSUE_3/IJARET_12_03_062.pdf
[4]. https://ymerdigital.com/uploads/YMER211037.pdf
[5]. https://ijcrt.org/papers/IJCRT2303922. pdf
[6]. https://ijsrset.com/paper/8152.pdf
[7]. Utilizing Machine Learning for recommending systems The assistant professor of computer engineering at SCOE in Kharghar, India is Suhasini Parvatikar.
[8]. T. Venkat Narayana Rao, Anjum Unisa, KothaSreni TERNATIONAL JOURNAL OF THE SCIENTIFIC TECHNOLOGY RESEARCHVOLUME 9, ISSUE 02, FEBRUARY 2020.