Grenze International Journal of Engineering and Technology, June Issue

Human Disease Prediction and Drug Recommendation

Varsha Bodade[1], Himani Bhole[2], Arya Sharma[3], Sudhir Rai[4] and Lav Bhangale[5]
[1]Professor, Terna Engineering College, Nerul Navi Mumbai, Maharashtra 400706, India
Email id varshabodade@ternaengg.ac.in
[2-5]UG Scholars, Terna Engineering College, Nerul Navi Mumbai, Maharashtra 400706, India
Email id himanivijaybhole@ternaengg.ac.in, aryasharma@ternaengg.ac.in

***Abstract*— The paper brings forward a novel point of view to human disease prediction along with drug recommendation utilizing machine learning (ML) techniques. With the increasing availability of healthcare data, there is a growing need for efficient methods to predict diseases accurately and recommend suitable treatments. Our study addresses this challenge by leveraging ML algorithms to analyse patient symptoms, medical history, and demographic information to predict the likelihood of various diseases. Additionally, our research aim to give personalized drug recommendation to patients on their predicted diseases. By integrating ML models with healthcare data, our system can offer tailored treatment plans that consider individual patient characteristics and medical histories. We propose a drug recommendation system that takes into account multi-disease scenarios, providing accurate drug recommendations to healthcare professionals. Our approach not only enhances the efficiency of disease diagnosis and treatment selection but also contributes to improved patient outcomes and healthcare delivery. By harnessing the power of ML, our research offers a promising solution to the complex challenges in healthcare.**

***Index Terms*— Machine Learning, Disease Prediction, Drug Recommendation, Decision Tree, Random Forest, NB Gaussian, KNN, Weighted KNN, Disease Classification Program.**

I. INTRODUCTION

These years, integration of machine learning (ML) techniques into healthcare has revolutionized disease prediction and treatment recommendation processes. The ability of ML algorithm to analyse huge amounts of patient data and recognise complex patterns has empowered the development of personalized medicine, leading to better patient results. Human disease prediction and drug recommendation using ML have garnered significant attention due to their potential to enhance healthcare delivery and reduce medical errors. This research paper aim to explore the intersection of ML and healthcare by investigating the application of ML algorithms in predicting human diseases and recommending suitable treatment options. By leveraging ML models, healthcare professionals can analyse diverse datasets encompassing patient symptoms, medical history, genetic information, and demographic factors to predict the likelihood of various diseases with high accuracy.

Furthermore, the utilization of ML in drug recommendation facilitates the selection of personalized treatment plans tailored to individual patient characteristics. By considering factors such as disease severity, drug efficacy, potential side effects, and patient preferences, ML-based drug recommendation systems can assist healthcare providers in making informed decisions to optimize patient care.

Grenze ID: 01.GIJET.10.2.212

II. LITERATURE REVIEW

The author presents a proposes a machine learning-based system. This system employs user-input symptoms for disease prediction and drug recommendation, utilizing algorithms like Naive Bayes (NB) and Random Forest, each with varying accuracy levels. Following disease prediction, the system suggests suitable medication. To enhance system capabilities, the paper proposes a dual-functionality system for disease prediction and medication recommendation. An interactive front-end interface is developed using Flask to facilitate user engagement with symptoms [1].The proposed methodology of the research emphasizes explores Disease detection using machine learning, highlighting the importance of early diagnosis. The objective is to create a predictive framework capable of accurately detecting potential diseases from patient symptoms. Leveraging machine learning algorithms, the model processes symptom data, enabling swift and precise medical condition prognosis. The research employs a vast dataset encompassing person under care's symptoms and leveraging medical data for training and validate the model, demonstrating its notable accuracy in disease prediction. Findings indicate the potential utility of the proposed model in facilitating early disease diagnosis and treatment, thereby enhancing healthcare outcomes. This research offers significant advancements in healthcare informatics, potentially impacting disease prevention, treatment, and management strategies [2].

The author proposed a Machine Learning (ML)-based for early Disease Prediction, The objective of the project is to create an Android application utilizing data mining algorithms, implemented with Python and Java. The app receives user symptoms and employs cloud-based Machine Learning to identify potential diseases. Raw user data is collected, allowing patients to compare their diagnosis with information available on healthcare websites. Additionally, patients can engage in online support group chats to exchange experiences with others suffering from similar ailments. The system matches user-provided symptoms with those stored in a database, utilizing a range of data analysis methods and a sophisticated learning algorithm to enhance accuracy [3].) In this study, a novel machine learning framework is proposed for various ailments and drug recommendation, aiming to provide precise medication suggestions for patients with diverse health conditions. The framework offers tailored recommendations for ailments such as cardiac issues, the common cold, fever, obesity, optical disorders, and orthopedic problems. Various supervised machine learning algorithms including Support Vector Machine (SVM), Random Forest, Decision Tree, and K-nearest neighbors are employed to generate these recommendations. The research conducted experimentation and evaluation on a dedicated sample dataset solely created for testing purposes, ensuring no reliance on medical practitioner data. Results indicate that the Random Forest classifier achieves a commendable recommendation accuracy of 96.87%, surpassing other classifiers tested. Consequently, this approach emerges as a promising solution for delivering reliable medication recommendations to patients within the healthcare sector [4]. They have developed a disease prediction system employing various machine learning (ML) algorithms. The dataset encompasses over 230 diseases for analysis. Utilizing symptomatology, age, and gender data, our system accurately predicts potential illnesses for individuals. Notably, the weighted K-nearest neighbors (KNN) algorithm outperformed other algorithms, achieving a prediction accuracy of 93.5%. Our model serves as a diagnosis tool, akin to a medical professional, facilitating early disease detection to ensure timely treatment interventions [5].

The research focuses system utilizing three data mining algorithms: Decision Tree Classifier, Random Forest Classifier, and Naive Bayes Classifier. Initially, each classifier is trained separately, and the results are analyzed. To ensure accurate predictions crucial for proper treatment, different prediction levels are assigned based on the outcomes from multiple classifiers. Upon analysis, it's observed that in many cases, all three classifiers predict the same diseases, while some diseases are predicted by two classifiers but not the third. In instances where all classifiers predict different ailments, the final prediction relies on the Naïve Bayes classifier due to its higher accuracy and absence of overfitting issues. This approach ensures a robust prediction framework for effective disease diagnosis and treatment planning [6].The project, disease prediction is based on patient symptoms input, allowing the system to accurately predict the disease. Classification algorithms such as Naïve Bayes (NB), Random Forest, Logistic Regression, and KNN are employed, each yielding varying accuracies. The system can predict specific diseases like Heart Disease or Diabetes by indicating "True" or "False" based on whether the individual has the disease. This predictive capability holds significant potential for future medical treatment. After disease prediction, the system recommends the type of doctor to consult. The project also includes an interactive frontend interface for symptom interaction. The entire Developed with Django, the model runs on a dedicated Django server enhancing accessibility and usability [7].

The study shows advancements in technology within the healthcare sector provide solutions by offering recommendations for doctors and hospitals, guiding patients on where to seek treatment and which specialists to consult. Their application implements this functionality by predicting suitable doctors and hospitals based on

feedback from patients and guardians, aggregating positive and negative reviews to assist in decision-making [8]. In this research, the author introduces the proposal advocates for utilizing mobile devices as gateways for transmitting patient health data, such as temperature, heart rate, and ECG readings, to a central medical server, designated caregivers, and healthcare providers. This data stream would be generated by sensors embedded in the mobile devices or deployed in the environment [9]. This paper aims to prevent emergencies by prompting patients with pre-emptive actions before their health necessitates urgent medical intervention. The system leverages a programmable ARM7 processor to analyse bio-signals and assess heart health. Upon detecting abnormalities in heart parameters, it promptly alerts the healthcare provider. The system transmits the ECG signal wirelessly via Bluetooth on the Android platform to a central medical server accessible by the doctor's PC. This allows for visual ECG evaluation on the server. For critical situations, additional alerts are sent to the doctor's mobile phone, ensuring notification even when away from the server. Testing confirms the device's compactness, cost-effectiveness, user-friendliness, and potential benefits. [10].

III. METHODOLOGY

A. Prediction Using Machine Learning

This section demonstrates the steps taken to analyse, predict disease, and recommend drug. The following actions were taken:

1) Data Acquisition

A dataset containing various disease cases is selected. The dataset should include relevant attributes like symptoms, age, and gender for model training.

2) Data Preprocessing

Real-world data often contains inconsistencies, errors, and missing values. Preprocessing addresses these issues. This stage involves techniques like data cleaning (correcting errors) and feature selection (removing irrelevant or redundant features).

3) Model Building

Different algorithms are employed to build models for disease prediction and drug recommendation. Common choices include Decision Trees, Random Forests, Naive Bayes Gaussian, K-Nearest Neighbors (KNN), and Weighted KNN.

4) Model Training and Evaluation

The preprocessed data is split into training and testing sets. The training set is used to train the chosen models, while the testing set is used to evaluate their performance (accuracy) in predicting diseases and recommending drugs. The model with the highest accuracy is selected for deployment.

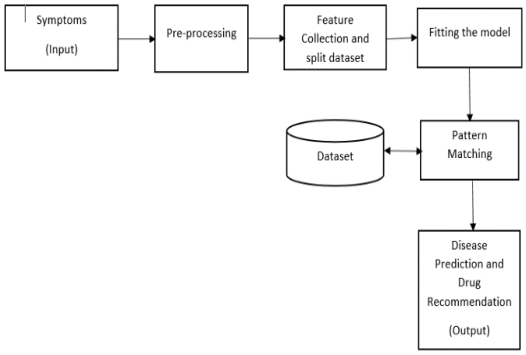

Fig. 1 Proposed Model Workflow

1) SVM -SVMs (Support Vector Machines) are a powerful tool commonly used in supervised learning tasks like disease prediction. They excel at finding patterns and decision boundaries in labelled datasets, where each data point has a known disease classification. However, for scenarios involving unlabelled data (data points without disease labels), SVMs cannot be directly applied for disease prediction. Here are some alternative approaches for leveraging unlabelled data in disease prediction.

Semi-supervised Learning: This technique incorporates a small amount of labelled data along with a larger set of unlabelled data to train the model. The labelled data guides the learning process, while the unlabelled data helps refine the model's ability to identify patterns.
Clustering Algorithms: These algorithms can be used to group unlabelled data points into potential disease clusters based on inherent similarities. This can be a preliminary step, and the resulting clusters can then be labelled by medical experts for supervised learning tasks like disease prediction using SVMs or other algorithms.
2) Decision Tree- Decision trees can be employed to classify patients into different disease categories based on their symptoms and medical history. Each node in the tree represents a symptom or attribute, and the branches represent the decision rules based on those attributes. It can help identify the most relevant features for disease prediction.
3) Naive Bayes- Naive Bayes algorithms, particularly Gaussian Naive Bayes, can estimate the probability of a patient having a particular disease given their symptoms. It assumes that features are independent of each other, which simplifies the calculation of probabilities.
4) KNN- KNN algorithms classify patients by finding the k nearest neighbors in the feature space and predicting based on the majority class among them. It can be useful for identifying similar patients and recommending suitable treatments or medications based on their similarities.
5) Weighted KNN- Weighted KNN assigns different weights to the nearest neighbors based on their distance from the query point. This approach can improve the accuracy of predictions by giving more importance to closer neighbors, thereby enhancing the relevance of drug recommendations.
6) Random Forest- Random forests, which are ensembles of decision trees, can provide more robust and accurate predictions by combining multiple decision trees. They can handle larger datasets and reduce overfitting, thus improving the accuracy of disease prediction and drug recommendation.

Working

- *Distance Calculation*: For a new instance to be classified or predicted, distances to all instances in the training set are computed. Common distance metrics include Euclidean distance, Manhattan distance, or Minkowski distance.
- *Nearest Neighbors*: The k instances with the smallest distances to the new instance are identified. These instances are referred to as the "nearest neighbors."
- In classification tasks, the new data point is assigned the class label that dominates its k nearest neighbors. For regression problems, the predicted value is the average of the target values from the k nearest neighbors.

B. Web Development

The Python-based backend uses scikit-learn for machine learning tasks, while the Flask and Angular in frontend offers an interactive interface for users to assess the likelihood of Disease Prediction and Drug Recommendation.

IV. Result & Discussion

Despite the limited data available for our research, our study demonstrates remarkable efficiency. We carefully considered factors such as the number of data points and features, data format, and employed a diverse set of at least five algorithms to optimize performance. We prioritized interpretability, feature selection, data point quantity, data linearity, and training time. Our comparative analysis reaffirms that our approach outperforms others in terms of efficiency.

A. Model Efficiency

Five algorithms were utilized to select the most accurate one for our research.

Model	**Training Accuracy**	**Validation Accuracy**
Decision Tree	1.0	1.0
Random Forest	1.0	1.0
Naïve Bayes	1.0	1.0
KNN	1.0	1.0
Weighted KNN	1.0	1.0

V. IMPLEMENTATION RESULT

Fig. 2 Sign in page

Fig. 3 Dashboard

Fig. 4 Symptoms

Fig. 5 Choosing Symptoms

Fig. 6 Adding Symptoms

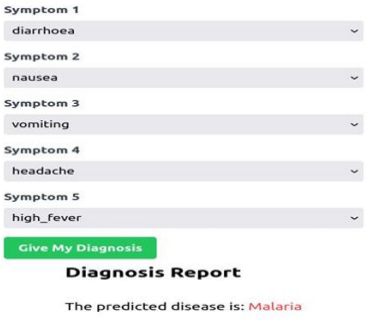

Fig. 7 Diagnosis Report

Fig.8 Drug Recommendation

Our research explored various disease prediction algorithms. The proposed model achieved a competitive accuracy rate compared to existing studies referenced in our work. While the model demonstrates promising performance, incorporating a larger dataset for training and testing could potentially enhance its efficiency and generalizability

VI. Conclusion And Future Work

In conclusion, tackling disease prediction and drug recommendation presents significant opportunities for early detection, personalized interventions, and enhanced outcomes. If we put symptoms like diarrhoea, nausea, vomiting, headache, high fever we get predicted disease as “Malaria”. In this experiment we found that KNN gives results superfast. Machine learning (ML) algorithms offer the capability to analyse intricate patterns and correlations within extensive datasets, aiding in the identification of risk factors, anticipation of potential cases, and customization of interventions to meet individual requirements. Through the integration of ML-powered solutions into research and clinical practices concerning diseases and drugs, we can pinpoint timely and tailored treatments for patients, thereby improving overall healthcare delivery.

References

[1] J., Siddegowda & Devi, A.. (2022). A Literature Review on Prediction of Chronic Diseases using Machine Learning Techniques. International Journal of Management, Technology, and Social Sciences. 28-49. 10.47992/IJMTS.2581.6012.0209.

[2] "Disease Prediction and Treatment Recommendation Using Machine Learning" by Saurav Mahata, Yash Bhaveshbhai Kapadiya, Vishal Kushwaha, Vatsal Joshi, Yassir Farooqui (2021)

[3] Mudaliar, Vivek, P. Savaridaasan, and Sachin Garg. "Disease prediction and drug recommendation android application using data mining (virtual doctor)." International Journal of Recent Technology and Engineering 8 (2019). International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-3, September 2019

[4] "A Drug Recommendation System for Multi-disease in Health Care Using Machine Learning" by Komal Kumar Napa and D.Vigneshwari (2020) Napa, Komal Kumar & Vigneswari, D. (2020). A Drug Recommendation System for Multi-disease in Health Care Using Machine Learning. 10.1007/978-981-15-5341-7_1.

[5] Keniya, Rinkal and Khakharia, Aman and Shah, Vruddhi and Gada, Vrushabh and Manjalkar, Ruchi and Thaker, Tirth and Warang, Mahesh and Mehendale, Ninad, Disease Prediction From Various Symptoms Using Machine Learning (July 27, 2020). Available at SSRN: https://ssrn.com/abstract=3661426 or http://dx.doi.org/10.2139/ssrn.3661426

[6] Gupta, Jay Prakash, Ashutosh Singh, and Ravi Kant Kumar. "A computer-based disease prediction and medicine recommendation system using machine learning approach." Int J Adv Res Eng Technol (IJARET) 12.3 (2021): 673-683.

[7] "Disease Prediction and Doctor Recommendation System using Machine Learning Approaches" [2021].

[8] “Disease Prediction and Hospital Recommendation using Machine Learning Algorithms” [2020]

[9] Nasiri, Mahdi, Behrouz Minaei, and Amir Kiani. "Dynamic recommendation: Disease prediction and prevention using recommender system." International Journal of Basic Science in Medicine 1.1 (2016): 13-17.

[10] Wahane Varsha, and P. V. Ingole. "An android based wireless ecg monitoring system for cardiac arrhythmia." 2016 IEEE Healthcare Innovation Point-Of-Care Technologies Conference (HIPOCT). IEEE, 2016.

[11] Dr. Varsha Bodade, Anuj Jaiswal, Abhishek kumar Ashok kumar Mishra, Lokesh Chaudhari ‘Health Record Management System”

[12] Nasiri, Mahdi, Behrouz Minaei, and Amir Kiani. "Dynamic recommendation: Disease prediction and prevention using recommender system." International Journal of Basic Science in Medicine 1.1 (2016): 13-17.