MedXpert: Disease Prediction and Drug Recommendation Using Machine Learning: A Review

Swati Mhase, Shivprasad Sakhare, Pooja Salbande, Dipak Sangale, Dr.Gunjal Sanjay
Department of Computer Engineering Sanjivani College of Engineering Kopargaon, A.Nagar, India
swatimhase2531@gmail.com, shivprasadsakhare19@gmail.com, poojasalbande2019@gmail.com, sangaledipak5435@gmail.com.gunjalsanjay@sanjivani.org.in

***Abstract*—** The focus in healthcare settings is disease prediction, which is utilized by data mining techniques, particularly machine learning algorithms. The objective is to create a classifier system to assist physicians in early disease prediction and diagnosis. Machine learning algorithms, including the Decision Tree classifier, Random Forest classifier, and KNN, are applied, and contrasted for their efficacy in disease prediction. The study underscores the potential impact of these algorithms on enhancing healthcare outcomes.

Implementing data mining procedures to scrutinize extensive medical data, especially within hospital information systems (HIS), is the second scenario. The goal is to unveil noteworthy patterns and trends within diagnosis case data, with the ultimate aim of furnishing patients with personalized information about their symptoms, illnesses, and appropriate medications.

The strategy encompasses the use of four distinct prototypes for disease prediction. The Vader tool and sentiment analysis, grounded in Natural Language Processing (NLP), are employed to scrutinize patient reviews. Subsequently, to suggest medications based on ratings and conditions. The models and strategies are elucidated comprehensively throughout the paper.

The experimental results presented imply that the proposed methodologies and models could be advantageous for drug recommendations. The application of sentiment analysis and data mining on medical databases has the potential to assist individuals in comprehending their health conditions and identifying suitable treatments.

Keywords— Disease Prediction, Drug Recommendation, KNN algorithm, Random Forest algorithm, Decision Tree algorithm.

I. INTRODUCTION

Machine Learning is the field of computer science that focuses on improving a computer's performance by enabling it to learn from data and past experiences. This learning process enhances the system's output, making it a powerful tool for a wide range of applications. One significant application is disease prediction and drug recommendation, which has seen considerable advancements in recent years. The process of machine learning involves two crucial stages: training and testing. During the training phase, the computer system is exposed to examples and previous data, allowing it to analyse and learn from this information. It uses this knowledge to improve its performance. The testing phase assesses the system's ability to apply what it has learned to new, unseen data. This iterative process refines the system's predictions and recommendations over time.

Disease prediction based on a patient's symptoms and medical history is one of the best applications of machine learning. By analysing datasets of medical information, machine learning algorithms can identify patterns and correlations that may not be clear to human experts. This helps in early detection and accurate prediction of diseases, potentially saving lives. Furthermore, machine learning has made significant strides in drug recommendation. The pharmaceutical industry benefits from this technology by accelerating drug discovery and development processes. Machine learning algorithms can analyse chemical structures, biological data, and historical drug performance to recommend potential candidates for drug development. This not only reduces costs but also expedites the delivery of new, life-saving medications to the market.

The integration of Machine Learning (ML) algorithms into medical diagnosis and decision-making processes heralds both promise and complexity. ML's reliance on high-quality datasets and precise feature selection underscores the significance of meticulous data curation and algorithmic fine-tuning. Errors stemming from biased or inadequate data and suboptimal feature selection can potentially compromise diagnostic accuracy, risking patient safety. However, proponents like Ismaeel argue against the limitations of traditional statistical techniques and human intuition, emphasizing ML's potential to mitigate biases

and uncover hidden insights within the burgeoning expanse of electronic health data. By harnessing advanced computational methodologies, ML endeavours to alleviate the burdens on medical staff while bolstering patient survival rates through early disease detection and informed decision-making. Thus, while challenges persist in ML's implementation, its integration into healthcare systems holds the promise of transformative improvements in diagnostic precision and patient outcomes.

In recent years, there's been a rise in systems that suggest medications and treatments for illnesses. These systems aim to give personalized advice to patients and doctors for better health outcomes. A recent study looked at how using Artificial Intelligence (AI) and Machine Learning (ML) can help with this. These techniques help find the right treatments for different diseases.

II. LITERATURE REVIEW

In recent times, there's been a rise in systems that recommend medications and treatments for illnesses. These systems aim to give personalized advice to both patients and healthcare professionals to enhance health outcomes. In a recent review, the advantages and challenges of leveraging artificial intelligence (AI) and machine learning (ML) techniques to provide personalized recommendations for both patients and clinicians were scrutinized. These advanced algorithms have been increasingly employed in healthcare settings to identify and propose tailored treatments for various diseases, aiming to optimize patient outcomes.[1]. Clinical errors are unfortunately common nowadays. In China, over 200 thousand people and in the USA, over 100 thousand people are affected each year due to prescription mistakes. It's crucial for patients to have doctors who have broad knowledge about bacteria, antibiotics, and individual patient needs when prescribing medication. Selecting the most appropriate medication is crucial for patient safety and treatment efficacy. Patients require doctors who possess comprehensive and up-to-date knowledge about various aspects of healthcare, including bacteria, antibiotics, and individual patient conditions. This broad understanding enables doctors to make informed decisions and prescribe medications that are most suitable for each patient's unique needs. [4] The diseases prediction system utilizes three data mining algorithms: Decision Tree Classifier, Random Forest Classifier, and Naive Bayes Classifier. Initially, a list of diseases along with their associated symptoms has been compiled. Then, an analysis is conducted on medications and their compositions in relation to the identified disorders. This process aims to develop a predictive model that can accurately classify diseases based on their symptoms and suggest appropriate medications based on their compositions and effectiveness in treating specific disorders [6]. Collecting disease symptoms from reputable health websites on the internet provides a valuable dataset reflecting real- world experiences. This approach offers diversity and authenticity but requires careful validation to ensure accuracy and address privacy concerns.[3]

III.WHAT IS AUTOMATED HEALTHCARE

Automation is changing many industries, including healthcare. But healthcare has been slower to adopt automation compared to other fields. One reason is that the term "automation" has been misused by vendors. Also, not focusing on tasks that could benefit from automation has slowed things down.

Automation can make healthcare better by supporting new technology and making processes work faster and smarter, which ultimately helps patients. Automation in healthcare refers to the use of technology to streamline processes, reduce manual tasks, and improve efficiency in various aspects of healthcare delivery. This can include automating administrative tasks such as appointment scheduling and billing, as well as clinical tasks such as diagnostics and monitoring.

System Architecture:

The overall architecture of the proposed approach is divided into three phases. In the first phase, user-friendly user interface. The second phase applies feature extraction to predict disease and to maintain patients' health record. The third phase drug recommendation will happen. Figure shows the system structure of the proposed approach. Details of all phases are described as follows.

- User interface- System has a user interface, which is implemented using python Django framework. Where users can interact with systems and can use resources.

- Feature generation: The feature generation is takes place here. Our features are based on the symptoms entered by users a random forest algorithm and Gaussian naive bayes algorithm.

- Feature extraction: Feature extraction methods in scikit- learn are typically implemented as transformers, and they can be used to select, transform, or create new features from your raw data.
- Machine learning algorithms: Random Forest is a powerful ensemble learning algorithm commonly used in machine learning for both classification and regression tasks.

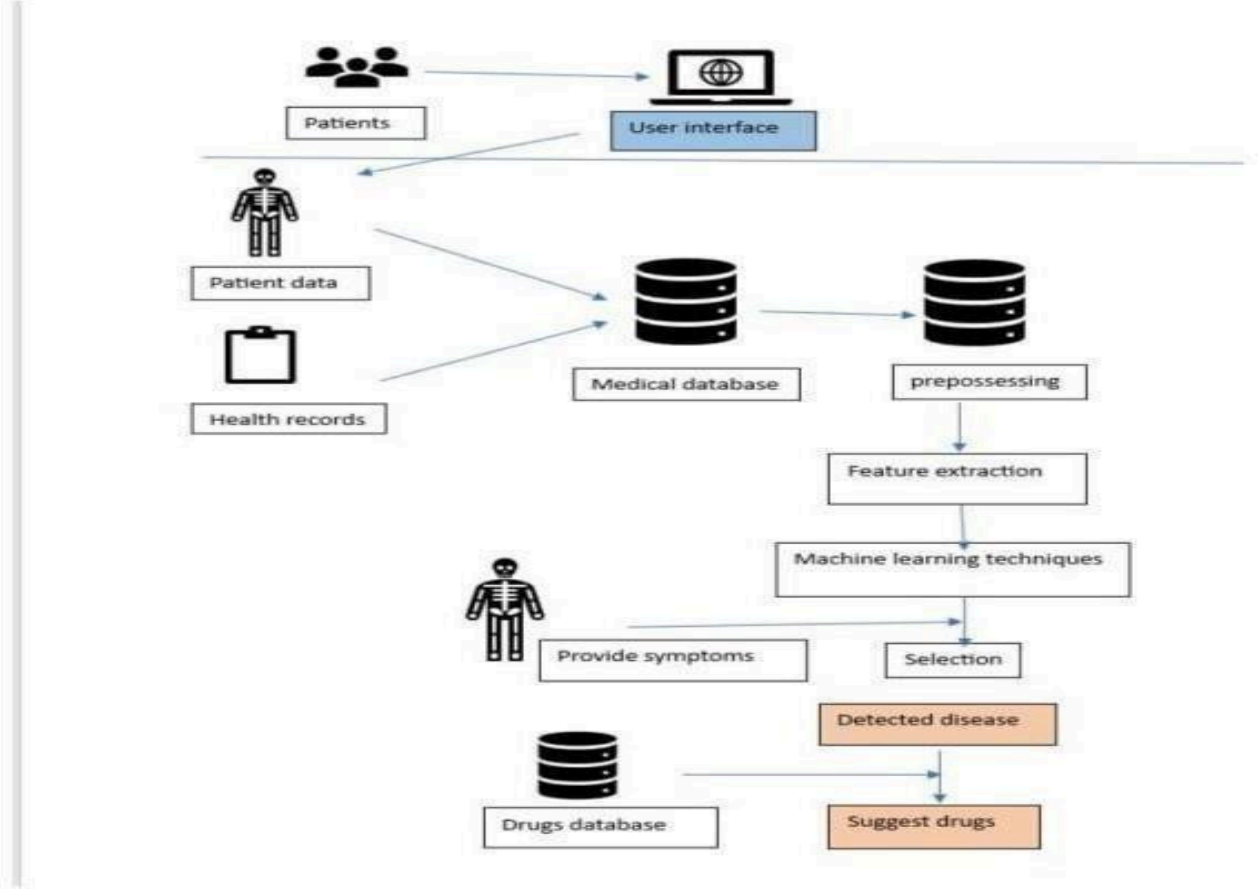

Figure 1: System Architecture

IV. METHODOLOGY

Method and analysis which is performed in your research work should be written in this section. A simple

1) KNN algorithm

The KNN algorithm is a non-parametric and instance-based supervised learning method used for both classification and regression tasks. It operates on the principle of similarity, where it predicts the class or value of a new data point by considering the majority class or averaging the values of its k nearest neighbors in the feature space.

The algorithm involves several steps:

- First, it identifies the k nearest neighbors of the new data point based on a chosen distance metric, commonly the Euclidean distance. Then, for classification tasks, it determines the majority class among the k neighbors, while for regression tasks, it calculates the average value of the target variable among the k neighbors.
- Parameter selection is critical in KNN, particularly choosing an appropriate value for k, which influences the algorithm's performance and generalization ability. Additionally, selecting a suitable distance metric is crucial, as it determines how the algorithm measures similarity between data points.
- Feature scaling is recommended in KNN to ensure that all features contribute equally to the distance calculation. Normalizing or standardizing features helps prevent variables with larger scales from dominating the distance computation.
- Once the parameters are set, KNN makes predictions for new data points based on their proximity to the training data. For classification, it assigns the class label with the highest frequency among the k nearest neighbors, while for regression, it calculates the mean or weighted average of the target variable among the k nearest neighbors.

1. Decision Tree

- A Decision Tree is a type of learning technique used in machine learning for solving problems like classifying data into categories or predicting outcomes. It's like a flow diagram where each step (node) represents a decision based on feature of the data.
- The Decision Node is where a decision is made based on a feature, like whether a patient has a fever.
- The branches represent the possible outcomes for that decision.
- The Leaf Node is the final outcome or category after following the branches from the decision nodes.
- Decision trees are often used for classification tasks because they're good at organizing and interpreting data to make decisions

2) Random Forest

An ensemble learning technique that calculates predictions from many decision trees to enhance accuracy. Random Forest is an ensemble learning algorithm widely used in machine learning for its ability to improve prediction accuracy. Random Forest constructs multiple decision trees forming a "forest", each trained on different dataset subsets and feature selections. This diversity reduces overfitting and enhances model robustness. During prediction, each tree offers its own prediction independently. In classification, the final prediction is decided by majority voting, while in regression, predictions are averaged.

In random forest, the following steps are conducted in the order listed:

- Step 1: Data pre-processing step.
- Step 2: Here Fitting the random forest algorithm to the training dataset.
- Step 3: Evaluate the result.
- Step 4: Now Test the accuracy of the result.
- Step 5: Visualize the result of the test group.

Random Forest is a powerful and versatile algorithm that is widely used in machine learning for various applications. It is particularly popular for its robustness and ability to handle complex and high-dimensional datasets. To use Random Forest in practice, you can leverage libraries like scikit-learn in Python, which provide easy-to-use implementations of the algorithm Gaussian naïve bayes:

The Gaussian Naive Bayes algorithm is a classification method rooted in probability theory, specifically Bayes' theorem. It's especially handy for tasks involving continuous or real-valued data. In simpler terms, it means that the algorithm treats each feature as independent of the others when calculating probabilities. Here's an overview of the Gaussian Naive Bayes algorithm.

Bayes' Theorem:

Bayes' theorem, it's a fundamental concept in probability theory, provides a way for calculating the probability of an event based on prior knowledge. Bayes' theorem can be expressed as:

$P(y|x) = P(x)$

$P(x|y) * P(y)$

$P(y|x)$ is the probability of class y given the features

x. $P(x|y)$ is the probability of observing features. x

given class P(y) is the prior probability of class y.

P(x) is the probability of observing features x.

V. RESULTS

The project's results encompass the accuracy and effectiveness of disease prediction and drug recommendation achieved using decision trees and other machine learning algorithms. The performance metrics include accuracy, precision, recall, and F1-score for disease prediction, along with recommendation accuracy, diversity, and coverage for drug recommendation. The interpretability of decision tree models and visualization of decision boundaries are highlighted. Additionally, comparisons with ensemble methods and validation in real-world healthcare settings demonstrate the prototype's efficacy in enhancing patient outcomes and personalized medicine.

Result Snapshot:

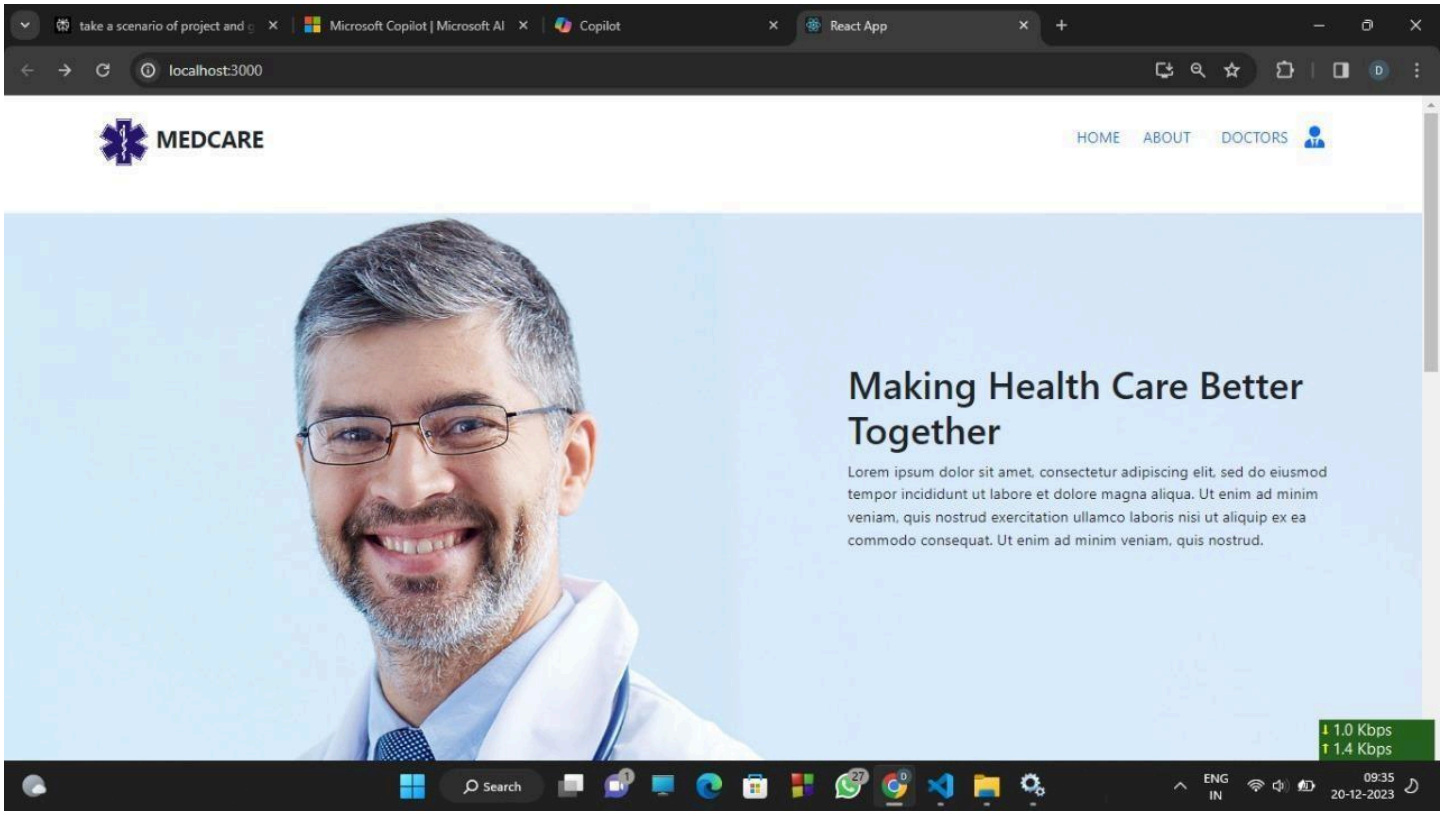

Figure 2: User Interface

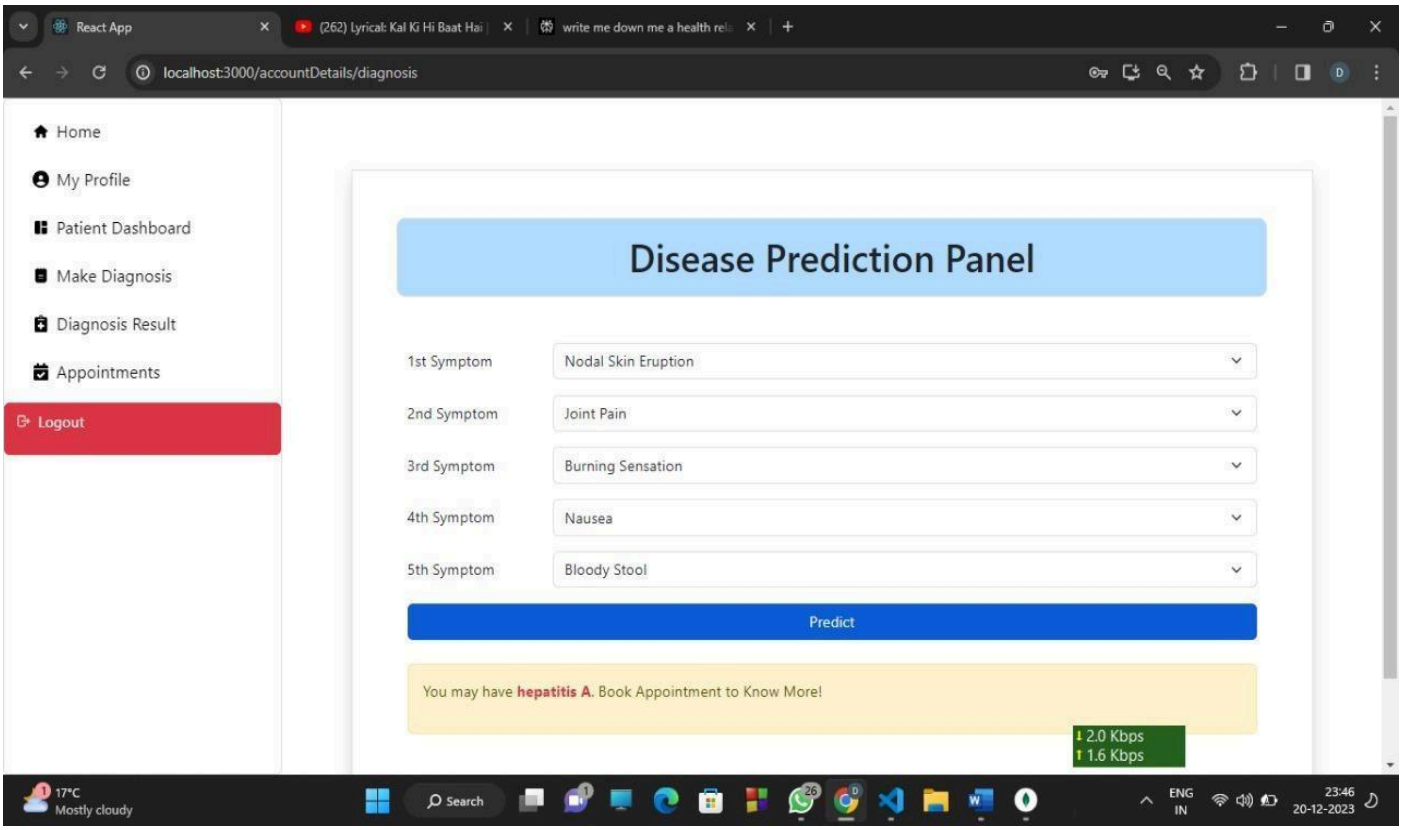

Figure 3: Patient Dashboard

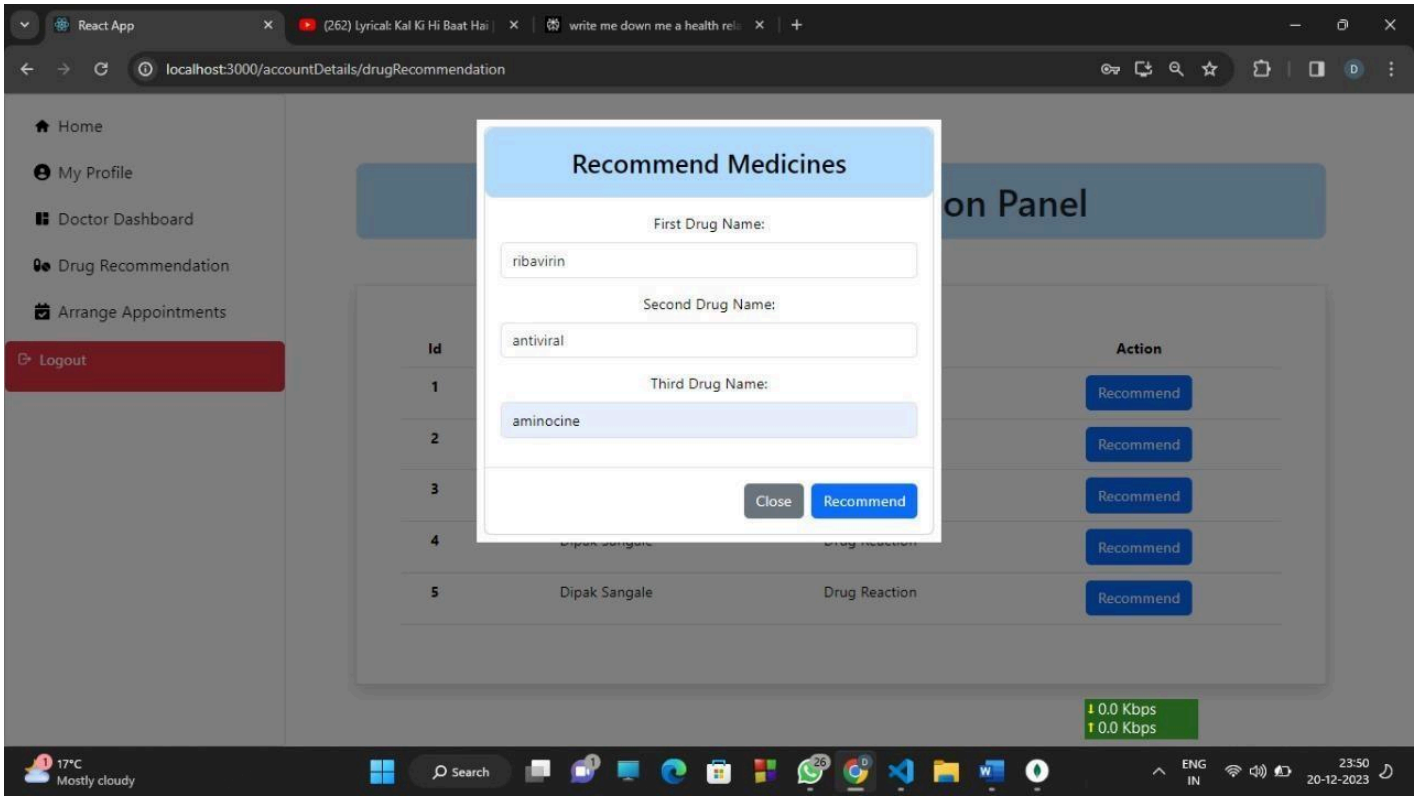

Figure 4: Doctor Dashboard

REFERENCES

[1] Nayak, S. K., Garanayak, M., Swain, S. K., Panda, S. K., & Godavarthi, D. (Year). "An Intelligent Disease Prediction And Drug Recommendation Prototype By Using Multiple Approaches Of Machine Learning Algorithms."

[2] Ferjani, M. F. (Year). "Disease Prediction Using Machine Learning." Bournemouth University, England.

[3] Reddy, P. P., Babu, D. M., Kumar, H., & Sharma, S. (Year). "Disease Prediction using Machine Learning." Lovely Professional University, Jalandhar, India.

[4] Garg, S. (Year). "Drug Recommendation System based on Sentiment Analysis of Drug Reviews using Machine Learning." Jaypee University of Information Technology, Solan, India.

[5] Isinkaye, O., Folajimi, Y. O., & Ojokoh, B. A. (Year). "Recommendation systems: Mining, Methods Appl., vol. 389, p. 408." Egyptian Informat. J., vol. 16, no. 3, pp. 261–273.

[6] Gupta, J. P., Singh, A., & Kumar, R. K. (Year). "A computer-based disease prediction and medicine recommendation system using machine learning approach." Int. J. Adv. Res. Eng. Technol. (IJARET), vol. 12, no. 3, pp. 673–683.

[7] Gupta, J. P., Singh, A., & Kumar, R. K. (Year). "A computer-based disease prediction and medicine recommendation system using machine learning approach." Int. J. Adv. Res. Eng. Technol. (IJARET), vol. 12, no. 3, pp. 673–683.

[8] Kononenko, I., Bratko, I., & Kukar, M. (Year). "Application of machine learning to medical diagnosis." Mach. Learn. Data.