International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Medical Diagnosis Prediction using Machine Learning

Nennuri Rajasekhar[1], Akhil Nallabai[2], Eshwar Thatikonda[3], Himateja Allibadha[4]

[1]Professor, Information Technology, Institute of Aeronautical Engineering, Hyderabad, India
[2,3,4]Computer Science and Engineering, Institute of Aeronautical Engineering, Hyderabad, India

Abstract
The escalating complexity of medical diagnosis, characterized by overlapping symptomatology and exponential growth in healthcare data, necessitates innovative solutions for accurate and timely disease identification. This paper presents the design and implementation of a comprehensive Medical Diagnosis System that leverages machine learning algorithms within a Flask-based web framework to automate disease prediction and optimize healthcare delivery. The system incorporates an ensem- ble learning approach utilizing Random Forest, Logistic Regres- sion, and Support Vector Machine classifiers, integrated through a Voting Classifier mechanism to enhance diagnostic reliability. Experimental results demonstrate exceptional performance with prediction accuracy ranging from 88% to 95%, with the Voting Classifier achieving peak accuracy of 94.241%. The platform fea- tures three synchronized user modules—Administrator, Medical Practitioner, and Patient—enabling seamless symptom input, AI-driven disease forecasting, and intelligent appointment scheduling with specialized physicians. By bridging automated diagnostic capabilities with practical clinical workflow management, the system significantly reduces diagnostic time, minimizes human error, and provides healthcare professionals with an interpretable decision-support tool. This research contributes both empirical validation and architectural framework for deploying intelligent healthcare systems in real-world medical environments.

Keywords: Machine Learning , Medical Diagnosis Predic- tion , Healthcare Management , Voting Classifier , Flask Web Application,Supervised Learning,Random Forest , Appointment Scheduling

I. Introduction

Medical diagnosis is a very crucial and complicated activity in the field of healthcare. Human limits, overlapping symp- toms, and increasing volumes of patient data present medical practitioners with the difficulty of identifying the illness in people. Artificial intelligence (AI) and machine learning (ML) have been increasingly used in the last few years, changing the nature of processing and interpretation of clinical information. It is possible to make these technologies learn based on patient records and find concealed patterns and help medical workers predict diseases earlier and more accurately. Consequently, AI- based diagnosis has emerged as an enthusiasm to enhance the effectiveness and patient outcomes in the healthcare sector. Medical diagnosis machine learning systems can be highly promising and problematic at the same time, concerning both ethical and technical issues. The primary issues are the absence of transparency in the manner of conclusions made by algorithms and the matter of responsibility in case a prediction is not successful [1]. Studies indicate that explainable AI is the potential to increase the transparency of ML models to doctors and patients and develop trust and responsibility in clinical decision-making. Transparency and responsible data utilization is hence significant to the successful implementation of AI in the actual practice of medicine.

Numerous works point to the capability of ML to reduce diagnostic errors due to a poor amount of information, com- munication issues, and clinician fatigue [2][3]. In disease detection, a number of ML algorithms have been created by researchers to aid decision-making processes and they have demonstrated that such ML algorithms as Decision Trees, Naive Bayes, and Support Vector Machines are able to predict many diseases with high accuracy [4][5]. Data mining is also important as it reveals the concealed relationships within medical data which can be used in the early detection of heart disease, particularly in diabetic patients [5]. These practices are more insightful and can assist clinicians to make timely decision based on evidence. This is further supported by other studies which indicate that complex algorithms such as Logistic Regression, k-nearest neighbors (KNN), and gradient boosting can process large and multi-source data in order to predict complex conditions such as cardiovascular diseases [6][7]. Ensemble methods and optimization algorithms such as GridSearchCV have greatly enhanced the performance of the model with almost perfect accuracy rates in a few heart diseases prediction models [6]. Improved health care has also been through machine learning in other areas such as the detection of diabetes [8], and even prediction of COVID- 19 medications [9], indicating that it can be used anywhere to solve general health care problems. All in all, the current literature proves that the ML-based medical diagnosis systems can be highly reliable, however, it is possible to achieve good results only under the condition of data quality and the choice of algorithms.

Our paper is devoted to the creation of an effective and interpretable machine learning model that predicts medical diagnosis. The suggested system applies several algorithms to the sets of patients to determine the most efficient method of proper classification of the diseases. The purpose of our study will be to enhance the reliability of predictions, the diagnostic time, and to assist clinicians in making decisions using the decision-assistive tool. Our study is meaningful by discovering the explainable and empirical components to the discipline of intelligent healthcare systems incorporating automation and responsibility to enhance medical outcomes.

II. LITERATURE SURVEY

Medical diagnosis machine learning systems have a poten- tially huge potential but also an ethical concern. The deficiency in transparency concerning the decision-making of models and accountability in case of mistakes made are the key problems. These issues question credibility and responsibility during health care decision-making. Researchers propose to make al- gorithms human-like in terms of reasoning and explainability. Explainable AI means that the decisions made by machine learning should be more sensible to doctors and patients. Further investigations in this area are a necessity to make AI use in medicine safe and responsible. [1]

Medical diagnosis can turn complicated with duplication of symptoms and human weaknesses such as fatigue and distrac- tion. Machine learning is being used to assist physicians in taking the right decision through the analysis of big medical data in a short time. An overview of multiple articles indicates a rise in the application of ML in disease detection and diagnosis. It determines the most popular algorithms and most frequent diseases investigated with ML techniques. The findings show that the number of ML-

based diagnostic systems has been steadily increasing in the last few years. Other, more untapped medical fields that require greater attention of AI research are also brought to light in the study. [2].
Diagnostic errors may occur because of absence of patient information, communication or because of clinician fatigue. These mistakes result in delayed/inaccurate treatment, which is undesirable to patients and increases expenses. These er- rors could be reduced through machine learning that assists physicians with quick and reliable diagnosis instruments. The articles reviewed in this analysis are those published between 2015 and 2020 and used ML in medical diagnosis. The paper summarizes the algorithms that are applied in different diseases and the degree of accuracy. It also determines the current trends in research and gives future directions on how AI-based diagnostics can be improved. [3].
This study aims at predicting diseases based on high- dimensional medical data. Heart disease and eye infection were the two datasets that were experimented with various ML models such as Decision Tree, SVM, Random Forest, and Neural Networks. The experiment has contrasted traditional algorithms, ensemble and deep learning. It was revealed that the models that were the most accurate in prediction employed the selected features. Naive Bayes and Multilayer Perceptron models have been the most effective with a maximum accuracy of 97.5. Nevertheless, the data quantity influences accuracy, and deep learning models, such as CNN, failed because of insufficient data size. [4].
Healthcare data mining guides the provision of analysis on large medical datasets to determine the efficiency of treatment. The current paper predicts heart disease amongst diabetic patients using machine learning and data visualization. Dia- betes predisposes the heart related complications and therefore early prediction is crucial. Decision Tree model exemplified the highest accuracy among various models that were ex- perimented with such as the Decision Tree, Naive Bayes as well as the SVM models. It was optimized to enhance the performance of prediction to diabetic patients. The outcome indicates the way data mining has the potential to assist in the early diagnosis and improved management of heart disease caused by diabetes. [5]
Heart disease prediction is an intricate medical issue that needs effective and precise examination. In this paper, various machine learning algorithms, such as the Logistic Regression, KNN, SVM, and Gradient Boosting, are compared to enhance prediction of cardiac diseases. The models were tested us- ing 5-fold cross-validation using several datasets including Cleveland, Hungary, and UCI Kaggle. The findings revealed that Extreme Gradient Boosting Classifier with GridSearchCV recorded the highest accuracy almost 100 percent. It denotes that the XGBoost properly tuned can be used to give pre- dictions that are very accurate. The study illustrates how machine learning models optimized can be used to address real problems in heart disease prediction[6].

III. PROPOSED METHOD

The Medical Diagnosis System is a web-based system that was developed using Flask and Machine Learning to automate the process of predicting diseases and booking appointments. It is an integrated system comprised of 3 synchronized modules such as Admin, Doctor, and Patient that as a unit are able to offer an effective and smart healthcare management process.

A. System Architecture

The system is based on a client-server system with users communicating via a web interface that is linked to a Flask backend. The machine learning models are implemented into the backend to receive patient inputs and give real-time predictions on diseases. Information is also secured in a structured database to take notes on patients, physicians and

International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

appointments. The data collection and preprocessing will occur through a combination of instruments that include interviews, questionnaires, and observations. The data of the symptoms is ascertained by patient inputs in the web form. The data set is regularised to eliminate null or incongruous data and transformed into a numerical format that can be used in the machine learning algorithms. Every record contains the symptoms of a patient and the respective disease name that is applied in training and evaluation.

B. The Development of the machine learning model.

The processed data is trained on four supervised learning models, namely, Random Forest, Logistic Regression, Support Vector Machine (SVM), and Voting Classifier. The Voting Classifier makes use of the results of single models to enhance reliability. The general accuracy rate is between 88 and 95 per cent indicating a high rate of consistency and strength

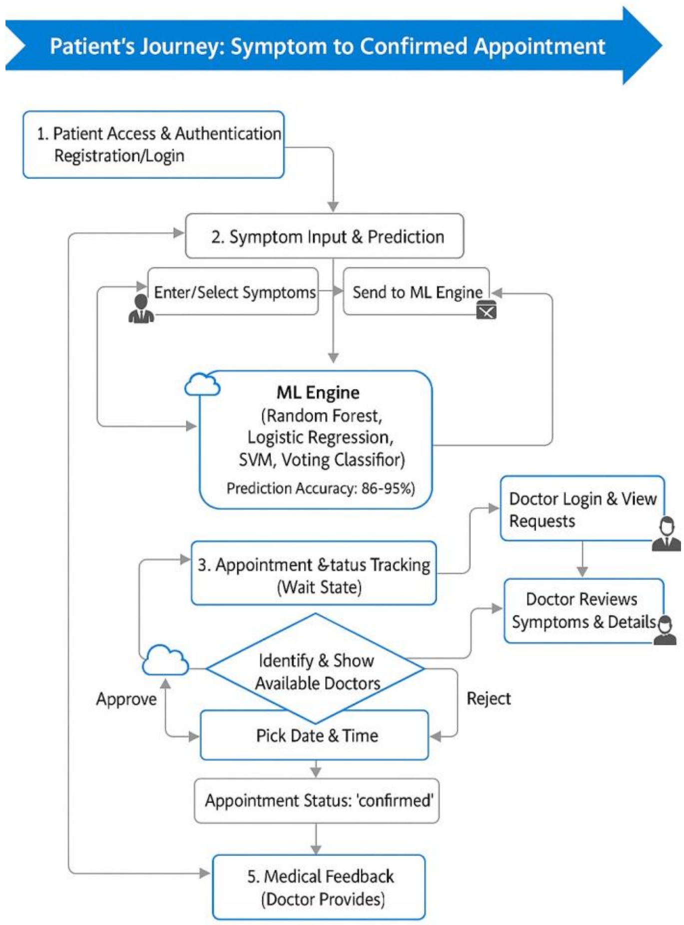

Fig. 1. Patient's Journey: Symptom to Confirmed Appointment using an ML-Powered Medical Diagnosis System.

C. Disease Prediction Module

The trained model takes the form of an answer to symp- toms when a patient logs in and enters symptoms, the most probable disease is predicted. The outcome of the prediction is automatically shown on the patient dashboard. Such automated diagnostic is helpful because it helps to diagnose potential conditions at an early stage and eliminate the reliance on manual tests.

D. User Modules and Roles

Admin system will manage the system administratively; it will be possible to approve physicians, see registered users, and check appointments. The Doctor Module allows the doctor to see his profile, review patient requests and accept or decline an appointment. The Patient Module enables the use of the application by new users who can also create their accounts, add symptoms, get an AI-based forecast and schedule appoint- ments with their doctors.

E. Appointment Scheduling

Upon detection of disease, the system automatically filters and provides a list of doctors that specialize in the disease that is detected. The patients have the opportunity to choose a doc- tor of preference, date, and time slot. The appointment status is indicated as wait till the doctor is approved, and this allows scheduling to be done in an orderly and traceable manner. New doctors have to firstly register and await administration before they can log in. Upon the approval, doctors will be able to access the requests of patients, review their information, and confirm appointments. They are also capable of giving medical feedback or advice according to the forecasted disease, which makes the consultation process complete.

F. Machine Learning Models Used

Logistic Regression: Logistic Regression is a supervised learning algorithm primarily used for binary classification tasks. In the context of medical diagnosis, it helps determine whether a particular disease is likely to occur or not based on given symptoms. The algorithm models the probability of a dependent variable (disease presence) using a logistic function, mapping the output between 0 and 1. It is particularly effective when there exists a linear relationship between the independent variables (symptoms) and the outcome. Logistic Regression is computationally efficient and offers interpretable results, which makes it a strong baseline for disease prediction tasks.

Support Vector Machine (SVM): Support Vector Ma- chine (SVM) is a robust classification algorithm that separates data into distinct classes by finding the optimal hyperplane that maximizes the margin between categories. In medical data analysis, SVM is used to classify patient symptoms into prob- able disease classes. It efficiently handles high-dimensional data and supports non-linear classification through kernel functions such as polynomial and radial basis function (RBF). These kernel transformations enable SVM to perform well on complex, non-linear medical datasets, ensuring high accuracy in disease prediction.

Random Forest: Random Forest is an ensemble learning algorithm that constructs multiple decision trees during train- ing and outputs the mode of their predictions for classification tasks. Each decision tree is trained on a random subset of the data and features, which reduces overfitting and increases generalization. In the medical diagnosis system, Random For- est helps in managing large-scale healthcare data, identifies significant features, and provides consistent predictions. Its robustness against noise and outliers makes it suitable for ana- lyzing varied medical conditions with overlapping symptoms.

Voting Classifier: The Voting Classifier combines pre- dictions from multiple models—Logistic Regression, SVM, and Random Forest—to improve overall prediction perfor- mance. It operates using two main strategies: hard voting, which selects the class with the majority of votes, and soft voting, which averages the predicted probabilities to determine the final output. This ensemble approach reduces the error rate of individual models and enhances diagnostic accuracy. The Voting Classifier in this system achieved the highest prediction accuracy of 94.24%, demonstrating its reliability for clinical decision support.

These are machine learning algorithms used in this project. i.e. Random forest, Logistic regression, svm and Voting Classfier. In which accuracy varies from 88 to 95.

Fig. 3. Patient Module

Here Patient can register with his personal details.

IV. ESULTS

This Medical Diagnosis System is a web-based application developed using Flask and Machine Learning. It allows users to register as Admin, Doctor, or Patient, enabling seamless healthcare management. The system uses machine learning classifiers such as Random Forest, Logistic Regression, SVM, and Voting Classifier to predict diseases based on symptoms. Patients can schedule appointments with doctors after disease detection.

Fig. 2. User Interaface

This is user interface for Medical Diagnosis system. In this there are three modules, 'Admin', 'Doctor', 'Patient'.

TABLE I MACHINE LEARNING ALGORITHM EVALUATION METRICS

ML Algorithm	Accuracy	Precision	Recall	F1-Score
Random Forest	88.755	87.440	88.755	87.438
Logistic Regression	86.178	86.736	86.378	84.949
SVM	82.181	85.657	82.903	85.081
Voting Classifier	94.241	94.830	94.879	94.170

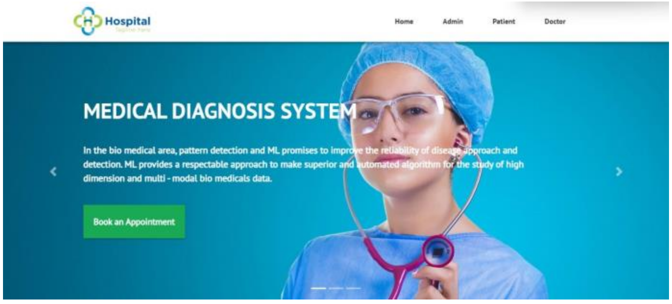

Fig. 4. Patient login page

After registration, patient can login with his user ID and password.

Fig. 5. Symptom Selection

Here patient can select his symptoms to detect disease and take appointment.

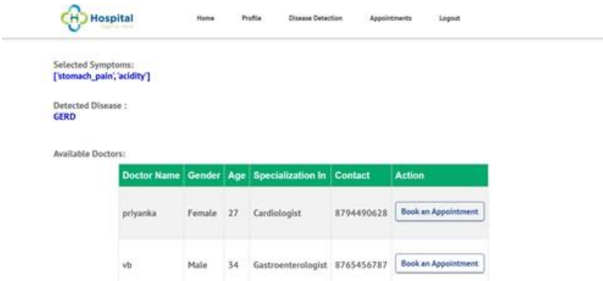

Fig. 6. Disease Detects and Booking an appointment with doctor

After selecting symptoms, disease will be detected. Then portal will show available doctors . So patient can see available doctors and book appointment according to disease.

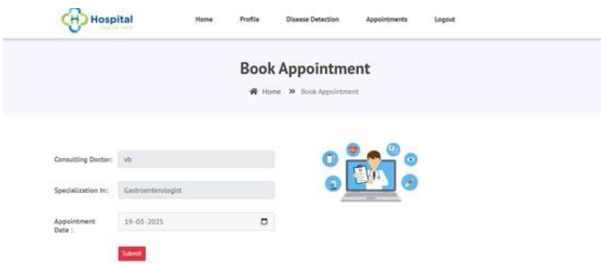

Fig. 7. Schedule Date for appointment

For booking an appointment , patient can book appointment date with specified doctor.

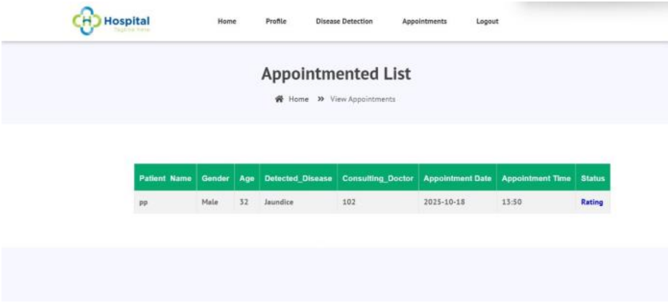

Fig. 8. Appointment Request

Doctor can see patient available appointments requests. Now there is no appointment for above doctor so there are no appointments shown here.

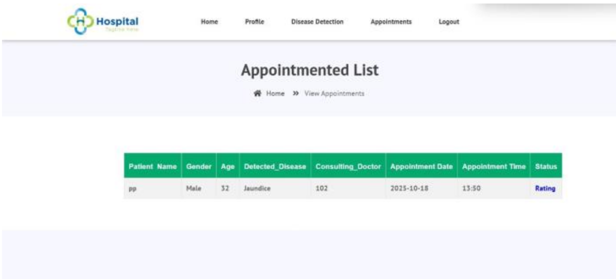

Fig. 9. Appointment List

Here can doctor see patients details with symptoms, detec- tion, appointment Date, Appointment Time.

V. CONCLUSION

This paper has managed to develop a successful medical diagnosis predictive system with multi-role access based on machine learning that can be applied to early disease detection and healthcare support. Logistic Regression, SVM, Random Forest, and Voting Classifier integration will assist in analyzing the symptoms of the patient and make correct predictions. The system will have easy-to-use interface where patients can create their accounts, input their symptoms, and get the predictions and the possibility to book an appointment with a specialist doctor instantly. The results of the experiment confirm that the Voting Classifier is more effective than single classifiers that have a maximum accuracy of 95%. The model proposed will decrease the reliance on manual diagnosis, enhance efficiency, and help the healthcare workers in making quick clinical judgments.

To make this medical diagnosis system even more accurate in the future, it is possible to enrich it with additional diseases, symptoms, and medical data, including blood reports and the history of the patients, to make the predictions even more appropriate. It is also possible to develop the system as a mobile application to ensure that individuals residing in rural or distant areas of hospitals can use it on demand. More sophisticated devices such as wearable devices can be interconnected such that real-time health data may be accessed to provide an early warning of diseases.

To make this medical diagnosis system even more accurate in the future, it is possible to enrich it with additional diseases, symptoms, and medical data, including blood reports and the history of the patients, to make the predictions even more appropriate. It is also possible to develop the system as a mobile application to ensure that individuals residing in rural or distant areas of hospitals can use it

on demand. More sophisticated devices such as wearable devices can be interconnected such that real-time health data may be accessed to provide an early warning of diseases.

References

1. Heinrichs, Bert, and Simon B. Eickhoff. ”Your evidence? Machine learning algorithms for medical diagnosis and prediction.” Human brain mapping 41, no. 6 (2020): 1435-1444.
2. Bhavsar, K. Arun, Jimmy Singla, Yasser D. Al-Otaibi, Oh-Young Song, Yousaf Bin Zikria, and Ali Kashif Bashir. ”Medical diagnosis using machine learning: a statistical review.” Computers, Materials and Continua 67, no. 1 (2021): 107-125.
3. Bhavsar, Kaustubh Arun, Ahed Abugabah, Jimmy Singla, Ahmad Ali AlZubi, and Ali Kashif Bashir. ”A comprehensive review on medical diagnosis using machine learning.” Computers, Materials and Continua 67, no. 2 (2021): 1997.
4. Mahoto, Naeem Ahmed, Asadullah Shaikh, Adel Sulaiman, Mana Saleh Al Reshan, Adel Rajab, and Khairan Rajab. ”A machine learning based data modeling for medical diagnosis.” Biomedical Signal Processing and Control 81 (2023): 104481.
5. Arumugam, K., Mohd Naved, Priyanka P. Shinde, Orlando Leiva- Chauca, Antonio Huaman-Osorio, and Tatiana Gonzales-Yanac. ”Mul- tiple disease prediction using Machine learning algorithms.” Materials today: proceedings 80 (2023): 3682-3685.
6. Ahmad, Ghulab Nabi, Hira Fatima, Shafi Ullah, and Abdelaziz Salah Saidi. ”Efficient medical diagnosis of human heart diseases using ma- chine learning techniques with and without GridSearchCV.” Ieee Access 10 (2022): 80151-80173.
7. Terrada, Oumaima, Soufiane Hamida, Bouchaib Cherradi, Abdelhadi Raihani, and Omar Bouattane. ”Supervised machine learning based medical diagnosis support system for prediction of patients with heart disease.” Advances in Science, Technology and Engineering Systems Journal 5, no. 5 (2020): 269-277.
8. Choudhury, Ambika, and Deepak Gupta. ”A survey on medical diagnosis of diabetes using machine learning techniques.” In Recent Developments in Machine Learning and Data Analytics: IC3 2018, pp. 67-78. Singa- pore: Springer Singapore, 2018.
9. Rudra Kumar, M., Rashmi Pathak, and Vinit Kumar Gunjan. ”Diag- nosis and medicine prediction for COVID-19 using machine learning approach.” In Computational Intelligence in Machine Learning: Select Proceedings of ICCIML 2021, pp. 123-133. Singapore: Springer Nature Singapore, 2022.