

HOUSE.AI - A MACHINE LEARNING BASED WEB APPLICATION FOR HUMAN DISEASE PREDICTION

¹Aakash. S, ²Aakash. B, ³Dharshan. S, ⁴Koushik. ZR, ⁵Ms. Lavanya. S

^{1,2,3,4}Student, ⁵Assistant Professor

Department of Computer Science and Engineering,
KCG College of Technology, Chennai, India

Abstract- House.AI is a web application designed to predict diseases using machine learning, with a focus on early detection. The application offers an intuitive interface for users to input their symptoms and receive predictions of potential illnesses. By training a machine learning model on a comprehensive medical record dataset, House.AI's algorithm makes accurate disease predictions based on user input. The application uses a combination of feature selection and extraction methods and machine learning algorithms to achieve high prediction accuracy. The House.AI application has the potential to aid healthcare professionals in diagnosing diseases early and assisting patients in taking necessary actions for their health.

Keywords- Web Application, Disease prediction, Machine Learning, Medical records.

I. INTRODUCTION

The emergence of machine learning algorithms in the healthcare industry has opened up new possibilities for early disease detection and prevention. With the vast amount of medical data available, machine learning algorithms can analyze and interpret these data sets to predict diseases with high accuracy. House.AI is a web application developed to predict diseases using machine learning. The application's primary objective is to provide a user-friendly interface that enables users to input their symptoms and receive disease predictions promptly. The application employs a combination of feature selection and extraction techniques and various machine learning algorithms to ensure accurate predictions. In addition to its disease prediction capabilities, House.AI has the potential to aid healthcare professionals in diagnosing diseases and help patients make informed decisions about their health. The application can also contribute to the early detection of diseases, which is critical for ensuring timely interventions and treatment. In this paper, we present a detailed overview of the House.AI application, including its objectives, methodology, and evaluation metrics. We also discuss the potential impact of House.AI on the healthcare industry and how it can assist medical professionals in delivering better care to patients.

Overview

House.AI is a web-based application developed to predict diseases using machine learning algorithms. The application is designed to be user-friendly, allowing users to input their symptoms and receive predictions of potential diseases promptly. The application employs the Django framework for server-side scripting, which provides a stable and scalable platform for the application. The random forest algorithm is used for the machine learning model, which is developed using Python and trained on large datasets of medical records to predict diseases with high accuracy. The user profile management system is handled using SQL databases, which ensures secure data storage and retrieval.

The frontend of the application is designed using HTML, CSS, and JavaScript, which provides a responsive and visually appealing interface for users. The user interface allows users to input their symptoms and receive disease predictions promptly. The application's machine learning algorithm utilizes feature selection and extraction techniques to obtain a subset of the most relevant features from the medical data. The selected features are then used to train the random forest model, which can predict diseases with high accuracy. The model's accuracy is continuously monitored, and the model is regularly retrained with the latest medical data to ensure the application's predictions are always up-to-date.

The House.AI application has several potential use cases, such as assisting medical professionals in early disease detection, providing timely interventions for patients, and helping patients make informed decisions about their health. The application's predictive capabilities can also be used to improve clinical trials, drug development, and disease research. Additionally, the application can be integrated with electronic health record (EHR) systems to provide a comprehensive medical diagnosis and treatment plan for patients.

In conclusion, House.AI is a disease prediction application that employs machine learning algorithms to assist healthcare professionals in early disease detection and provide timely interventions for patients. The application's use of Django for server-side scripting, random forest algorithm for machine learning, SQL databases for user profile management, and HTML, CSS, and JavaScript for frontend design makes it a robust and reliable tool for disease prediction. The application's user-friendly interface, accuracy of predictions, and potential impact on the healthcare industry make it a valuable addition to the field of medical research and healthcare.

Random Forest Algorithm

The random forest algorithm is an ensemble learning technique that enhances prediction accuracy and reliability by combining multiple decision trees. During the training phase, it constructs a collection of decision trees, with each tree trained on a random subset of the available features and data samples. Each decision tree independently generates an output, and the final prediction is determined through a combination or voting mechanism. This approach improves the robustness and effectiveness of the algorithm in producing accurate predictions.

One of the key advantages of the random forest algorithm is that it is resistant to overfitting, a common problem in machine learning where the model performs well on the training data but poorly on unseen data. The algorithm achieves this by randomly selecting features and data samples to train each tree, which helps to reduce correlation between the trees and improve their diversity.

Another advantage of the random forest algorithm is that it can handle high-dimensional feature spaces and datasets with missing values and noisy features. The algorithm is also relatively easy to use, and its computational overhead is low, making it suitable for large datasets and real-world applications.

The random forest algorithm has several applications, including classification and regression tasks in areas such as healthcare, finance, and marketing. The algorithm can also be used for feature selection and feature importance analysis, which can provide insights into the underlying data and improve the interpretability of the model.

To put it succinctly, the random forest algorithm is a widely used and effective machine learning method that enhances prediction accuracy and reliability by combining multiple decision trees. It overcomes the problem of overfitting and is capable of handling datasets with high dimensions, as well as missing or noisy features. Its low computational overhead and ease of use make it suitable for real-world applications, and its ability to provide insights into feature importance can improve the interpretability of the model.

Algorithm's objective

In House.AI, the primary goal of employing the random forest algorithm is to precisely determine the probability of a patient having a specific disease, utilizing their medical history and symptoms as input. This is accomplished by creating a collection of decision trees that are trained on a subset of the available features and data samples. By doing so, the algorithm aims to achieve accurate predictions regarding disease likelihood.

In House.AI, the primary aim of the random forest algorithm is to serve as a dependable and precise tool for healthcare professionals, aiding in the diagnosis and treatment of patients. By leveraging machine learning to analyze extensive datasets of patient information, the algorithm has the ability to detect hidden patterns and correlations that might elude human experts. This capability enhances the accuracy and speed of diagnosis, leading to improved patient outcomes. Moreover, the integration of machine learning algorithms in healthcare holds the potential to reduce healthcare expenses and optimize resource allocation, underscoring its growing significance in research and development.

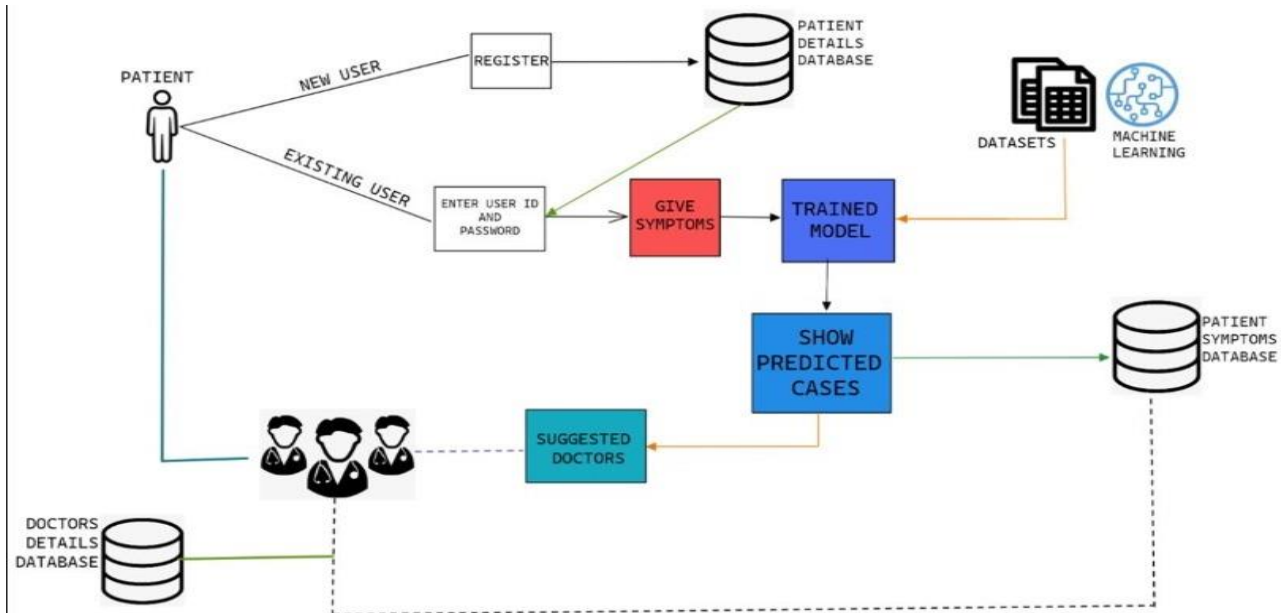
The features used by the algorithm to make predictions may include a patient's age, gender, medical history, symptoms, and other relevant factors. The algorithm considers each feature when constructing the decision trees, and the final prediction is based on the output of all the trees in the forest.

The random forest algorithm's objective is to improve the accuracy and robustness of the disease prediction by combining the predictions of multiple decision trees. By aggregating the outputs of several trees, the algorithm is less likely to overfit the data and more likely to make accurate predictions on unseen data.

The random forest algorithm's ability to handle high-dimensional datasets, missing values, and noisy features makes it an ideal choice for disease prediction applications. The algorithm is also relatively easy to use and has low computational overhead, making it suitable for real-world scenarios.

To summarize, the primary aim of the random forest algorithm in our disease prediction application is to precisely determine the probability of a patient having a specific disease by utilizing their medical history and symptoms. The algorithm achieves this by employing a group of decision trees that are trained on a subset of the available features and data samples. Its objective is to enhance the accuracy and reliability of predictions by combining the outputs of multiple trees.

II. ARCHITECTURE



House.AI is a comprehensive and reliable disease prediction application that utilizes machine learning to analyze large datasets of patient information. The system architecture consists of several modules, each responsible for a specific task. The User Management module provides user authentication, registration, and profile management, creating a personalized environment for users to access the application and view their disease prediction history. The Disease Prediction module utilizes the random forest algorithm to provide accurate and reliable predictions. The Doctor Details module provides a comprehensive directory of healthcare professionals, enabling users to search for doctors based on location, specialization, and other criteria. The Chat module provides a platform for users to communicate with doctors and healthcare professionals in real-time, allowing for quick and efficient diagnosis and treatment.

The Admin Management module enables administrators to manage user accounts, monitor system performance, and perform other administrative tasks, while the Feedback module allows users to provide feedback on the application's performance and suggest improvements. The combination of these modules ensures that the application is efficient, scalable, and easy to maintain, while providing users with a reliable and personalized disease prediction tool. Overall, the system architecture of House.AI is designed to provide a seamless and intuitive user experience while performing complex disease prediction tasks using machine learning.

III. MODULES

A) User Management Module

The User Management Module is responsible for user authentication and registration. It contains functionalities for user login, registration, and profile management. This module is also responsible for assigning different roles to different users, such as admin, doctor, or patient. The User Management Module ensures that only authorized users can access the application.

B) Disease Prediction Module

The Disease Prediction Module predicts the possible diseases based on the analysis of the symptoms entered by the patient. It has a database of diseases and symptoms and uses machine learning algorithms to match the input symptoms with the diseases and generate a list of possible diseases. The output is presented to the user along with the probability of each disease. The module also supports integration with external APIs for advanced disease diagnosis.

C) Doctor Details Module

The Doctor Details Module contains details of doctors, such as their name, specialty, availability, and location. It allows users to search for doctors based on their specialty, location, and availability. The module also allows users to

book appointments with doctors. The Doctor Details Module ensures that patients can easily find the right doctors for their specific needs.

D) Chat Module

The Chat Module provides a chat interface for patients to communicate with doctors. It includes functionalities like sending messages, receiving messages, and attaching files. It also contains features like notifications, alerts, and online status. The Chat Module ensures that patients can easily communicate with doctors and get answers to their queries in real-time.

E) Admin Management Module

The Admin Management Module is responsible for managing the overall application. It allows the admin to add and remove doctors, update their profiles, and manage the database of diseases and symptoms. The module also provides analytics and reports on the usage of the application. The Admin Management Module ensures that the application is managed efficiently and effectively.

F) Feedback Module

The Feedback Module collects feedback from the users about the application, doctors, and services. It includes functionalities like rating, reviews, and feedback submission. The Feedback Module ensures that the application continuously improves based on user feedback.

IV. FUTURE SCOPE

There are several potential areas of future development that could enhance its functionality and usefulness. For example, one potential area of development could be the integration of additional machine learning algorithms to further improve the accuracy of disease predictions. This could involve exploring deep learning algorithms or other advanced techniques to improve the precision and speed of disease prediction.

Expansion of User base and Multilingual Support : A potential area of development for House.AI could be the expansion of its user base beyond the local community. This could involve developing multilingual support for the application, making it accessible to users from diverse backgrounds and cultures. Additionally, the application could be further developed to include a wider range of diseases and conditions, allowing it to become a more comprehensive tool for disease prediction.

Advanced features of Doctor-Patient Communication and Health Data tracking : Another potential area of development for House.AI is the incorporation of more advanced features for doctor-patient communication. This could involve the integration of video conferencing, voice calling, and file sharing capabilities, enabling doctors and healthcare professionals to provide remote consultations and follow-up care. Additionally, the application could be further developed to include more advanced features for tracking and monitoring patient health data, providing doctors with a more complete picture of a patient's health and enabling them to make more informed decisions about treatment options.

Integration with wearables and other health monitoring devices: Another potential area of development could be to integrate House.AI with various wearables and health monitoring devices, such as smartwatches and fitness trackers. This would enable the application to track and analyze a wider range of health data, providing doctors and patients with a more complete picture of a patient's health.

Personalized recommendations and reminders: House.AI could be further developed to provide personalized recommendations and reminders for users based on their health data and medical history. For example, the application could provide reminders for medication schedules or recommend specific lifestyle changes based on a patient's health status.

Integration with electronic health records (EHRs): Another potential area of development could be to integrate House.AI with electronic health records (EHRs) used by healthcare providers. This would enable doctors and other healthcare professionals to access a patient's medical history and health data directly from the application, making it easier to provide personalized care and treatment recommendations.

Collaboration with research institutions: House.AI could collaborate with research institutions to gather more data and improve the accuracy of its machine learning algorithms. This could involve partnering with hospitals, universities, or other research organizations to collect and analyze data from a larger and more diverse patient population.

Integration with telemedicine platforms: House.AI could be integrated with telemedicine platforms to enable remote consultations and care. This would enable doctors and healthcare providers to provide care to patients who are unable to visit a clinic or hospital in person, improving access to healthcare for individuals in remote or underserved areas.

V. CONCLUSION

In conclusion, the House.AI application serves as a valuable tool for disease prediction, providing users with quick and accurate predictions based on machine learning algorithms. The system architecture of the application, which includes modules for user management, disease prediction, doctor details, chat, admin management, and feedback, is designed to provide a comprehensive and user-friendly experience.

Moreover, the use of the Random Forest algorithm in the machine learning model has demonstrated its effectiveness in disease prediction, with high accuracy rates and efficient processing times. The application's potential for future development in areas such as expanding the user base, incorporating advanced features for doctor-patient communication, and including a wider range of diseases and conditions, presents exciting opportunities for further research and innovation.

Overall, the House.AI application represents a significant contribution to the field of healthcare technology, with the potential to improve healthcare outcomes and enhance the patient experience.

REFERENCES:

1. Chen, Min, Yixue Hao, Kai Hwang, Lu Wang, and Lin Wang. "Disease prediction by machine learning over big data from healthcare communities." *Ieee Access* 5 (2017): 8869-8879.
2. Keniya, Rinkal, Aman Khakharia, Vruddhi Shah, Vrushabh Gada, Ruchi Manjalkar, Tirth Thaker, Mahesh Warang, and Ninad Mehendale. "Disease prediction from various symptoms using machine learning." *Available at SSRN 3661426* (2020).
3. Kosarkar, Nidhi, Pallavi Basuri, Poonam Karamore, Prachi Gawali, Pradnya Badole, and Pranjali Jumle. "Disease Prediction using Machine Learning." In *2022 10th International Conference on Emerging Trends in Engineering and Technology-Signal and Information Processing (ICETET-SIP-22)*, pp. 1-4. IEEE, 2022.
4. Kulkarni, Sayali, Isha Sawant, Megha Shinde, Vishal Sonar, and Vaishali Latke. "Django Website for Disease Prediction using Machine Learning." *Journal of University of Shanghai for Science and Technology* 23, no. 6 (2021): 1199-1205.
5. Kumar, N. Komal, and D. Vigneswari. "Hepatitis-infectious disease prediction using classification algorithms." *Research Journal of Pharmacy and Technology* 12, no. 8 (2019): 3720-3725.
6. Swathi, S., E. Saranya, R. M. Prabakaran, M. Sachin Kumar, and S. Bairavel. "Virtual Health Assistant." In *2021 International Conference on System, Computation, Automation and Networking (ICSCAN)*, pp. 1-4. IEEE, 2021.

A large, light blue watermark logo is centered on the page. It features a stylized figure with arms raised, holding a rectangular box with the letters 'IJRTI' in white. Below the box is a semi-circular shape. The logo is semi-transparent and serves as a background for the text.

IJRTI