WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Machine Learning-Based Framework for Disease Prediction and Doctor Recommendation

Prof. Priyanka Sheelavantar[1], Sahana Nimbalakar[2], Sneha Kashannavar[3], Akshata Gurai[4], Nivedita Latur[5]

[1,2,3,4,5]Department of Computer Science and Engineering, Angadi Institute of Technology and Management, Belagavi-590009, India

ABSTRACT:

The integration of machine learning techniques with web-based systems has shown significant promise in enhancing diagnostic accuracy and healthcare accessibility. This research presents a web-based disease prediction and consultation system developed using supervised machine learning algorithms and the Django framework. The system accepts user-input symptoms, processes them through trained models, and predicts the most probable disease. It further recommends a relevant doctor for consultation based on the diagnosed condition. The dataset used contains 4920 records across 132 symptoms and 41 diseases. After data preprocessing and feature selection, 95 significant symptoms were used for model training. Classification algorithms including Decision Tree, Random Forest, Gradient Boosting, K-Nearest Neighbors, and Gaussian Naïve Bayes were applied. Among them, the Decision Tree algorithm achieved the highest prediction accuracy of 97.51%. The system also provides a user-friendly web interface with role-based access for doctors and patients, enabling real-time communication and consultation. This study demonstrates the effectiveness of combining machine learning and web technologies in creating intelligent, scalable, and accessible healthcare solutions.

Keywords: Machine Learning, Disease Prediction, Web Application, Django, Doctor Recommendation, Decision Tree Classifier.

Introduction:

In the digital age, the integration of Artificial Intelligence (AI) and Machine Learning (ML) into healthcare systems is transforming traditional diagnostic practices. As the volume of patient data increases and disease symptoms become more complex, there is a growing need for intelligent systems that can support early diagnosis and treatment planning. ML offers powerful tools for analyzing patient data, identifying patterns, and predicting potential health risks based on symptom input, thereby enhancing both the accuracy and efficiency of disease diagnosis [1].

The availability of vast amounts of structured and unstructured medical data has enabled the development of predictive models to assist healthcare professionals in making informed decisions. Techniques such as Decision Tree, Random Forest, K-Nearest Neighbors (KNN), Gradient Boosting, and Naïve Bayes classifiers have demonstrated high accuracy in disease classification tasks [2][3]. When properly trained, these models can uncover complex relationships between symptoms and medical conditions, providing valuable support for clinical decision-making [4].

The utility of ML models is further amplified when integrated with real-time, user-friendly web applications. Web frameworks like Django enable the development of interactive platforms where patients can input symptoms and receive immediate diagnostic feedback. This functionality is particularly vital in regions with limited access to healthcare facilities, where early detection can significantly improve patient outcomes [5].

This study proposes a machine learning-based disease prediction and doctor recommendation system developed using the Django framework. The system allows users to input symptoms, which are processed by trained ML models to predict the most likely disease. Based on the prediction, the system recommends a suitable doctor for online consultation. The primary goal is to bridge the gap between patients and healthcare providers—especially in remote or underserved areas—through intelligent, technology-driven solutions.

The dataset used in this study consists of 4,920 patient records encompassing 132 distinct symptoms and 41 disease categories. After preprocessing and feature selection, 95 highly predictive symptoms were retained. Among the five algorithms implemented, the Decision Tree classifier achieved the highest accuracy, making it the most effective model for this application.
The rest of this paper is organized as follows: Section 2 outlines the objective and methodology, Section 3 details the system architecture and technologies used, Section 4 presents implementation and results, and Section 5 concludes with future work.

Methodology:

As healthcare applications adopt machine learning and real-time web technologies, disease prediction systems have become more accurate and accessible. This research integrates supervised machine learning models with a role-based web interface to predict diseases from symptoms and connect patients with relevant doctors. The methodology comprises data preparation, model development, system architecture, and deployment.

A. Disease Prediction Approach

Machine learning algorithms were employed to predict diseases based on user-input symptoms. The dataset used comprises **4,920 records**, each mapping **132 symptoms** to **41 diseases**, collected from publicly available health data repositories such as Kaggle [11].

After initial cleaning, **95 highly relevant symptoms** were selected through feature selection techniques to reduce dimensionality and avoid overfitting.

There are five supervised learning models used:

- Decision Tree Classifier
- Random Forest Classifier
- Gradient Boosting Classifier
- K-Nearest Neighbors (KNN)
- Gaussian Naïve Bayes

Each model was trained using the Scikit-learn library in Python and evaluated on both training and test datasets. The **Decision Tree** model achieved the highest training accuracy of **97.51%**, while others performed equally well on test data with **97.11%** accuracy.

B. Prediction Flow and Model Architecture

The disease prediction system accepts symptoms from the user and processes them using the trained machine learning model. The system then outputs the most probable disease along with a list of doctors for consultation.

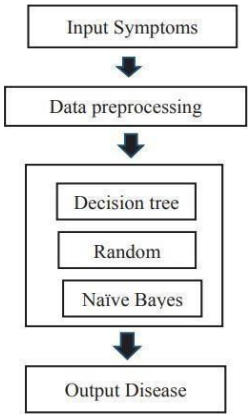


Fig. Prediction Model

C. System Implementation and Workflow

The web system was built using **Django**, a Python-based framework, and provides two user roles: **patient** and **doctor**. Patients input symptoms and receive predictions. Doctors view consultations and respond to patient queries.

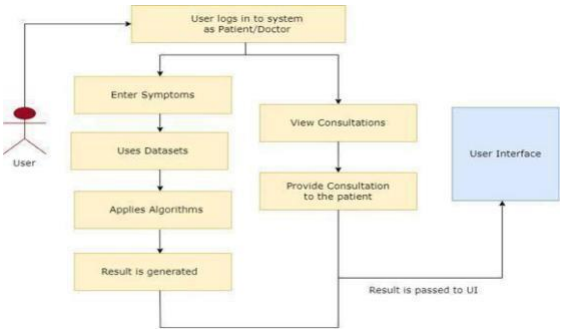


Fig. System implementation Flowchart

D. Technologies and Tools Used

To implement the system, the following technologies were used:

Component	Technology
ML Algorithms	Python, Scikit-learn
Web Framework	Django (Python)
Frontend	HTML, CSS, Bootstrap, jQuery
Backend Logic	Django Views, Flask, Jinja2
Database	PostgreSQL
Chat Feature	Django Channels

All user interactions are performed on a web interface with real-time communication features, session management, and data handling via a PostgreSQL database.

E. Summary of Workflow

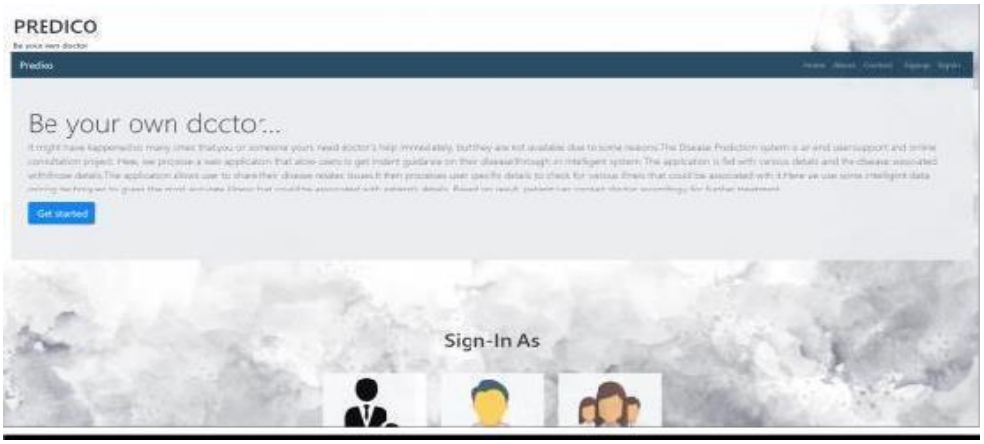
1. **User Registration/Login** – Role-based access for patient or doctor.
2. **Symptom Input (Patient)** – User enters observed symptoms.
3. **Prediction (Model)** – Backend ML model processes symptoms and predicts disease.
4. **Doctor Consultation** – System matches doctor specialization and initiates real-time communication.
5. **Doctor Interface** – Doctor can view patient data and respond

Objective:

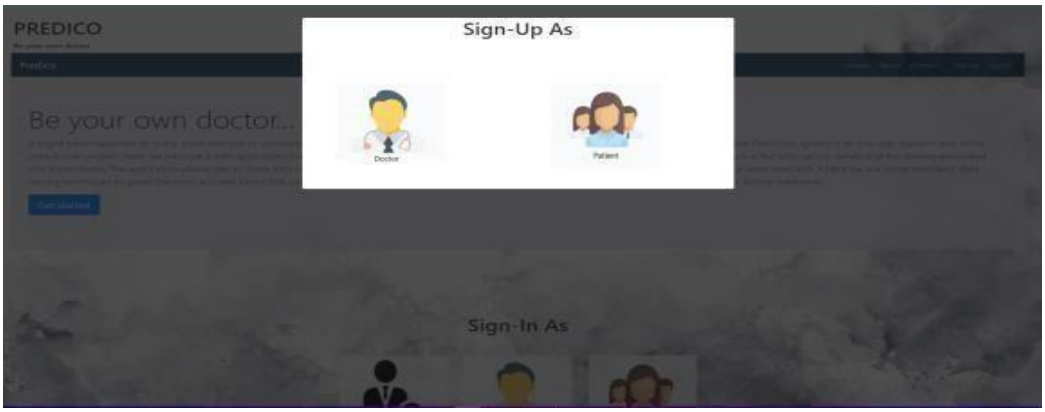
1. To research and compare various supervised machine learning algorithms for illness prediction.
2. To accurately predict specific diseases based on patient symptoms in the near term.
3. To develop an intelligent symptom-based disease prediction system that improves diagnosis accuracy by allowing users to input multiple symptoms.

Results

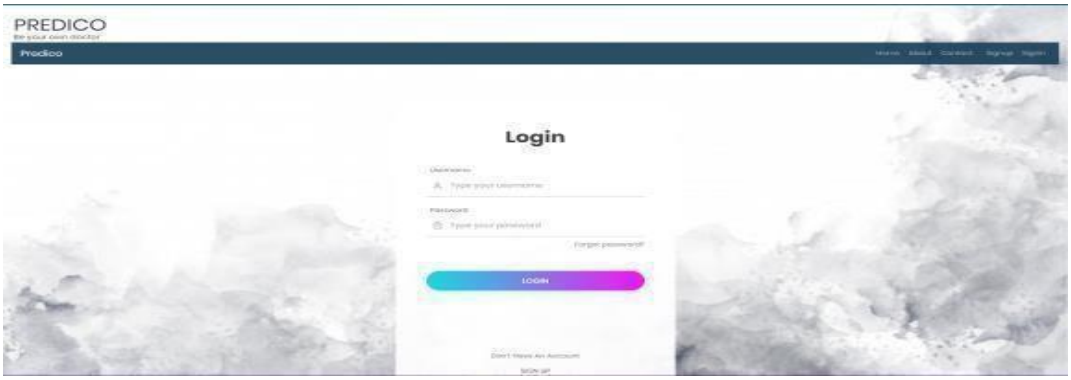
Home Page:



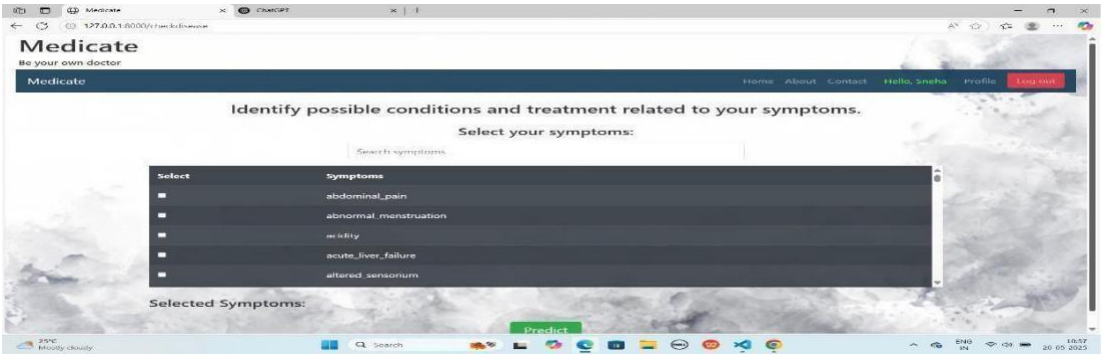
Sign-up Page :



Patient/Doctor Login Page:

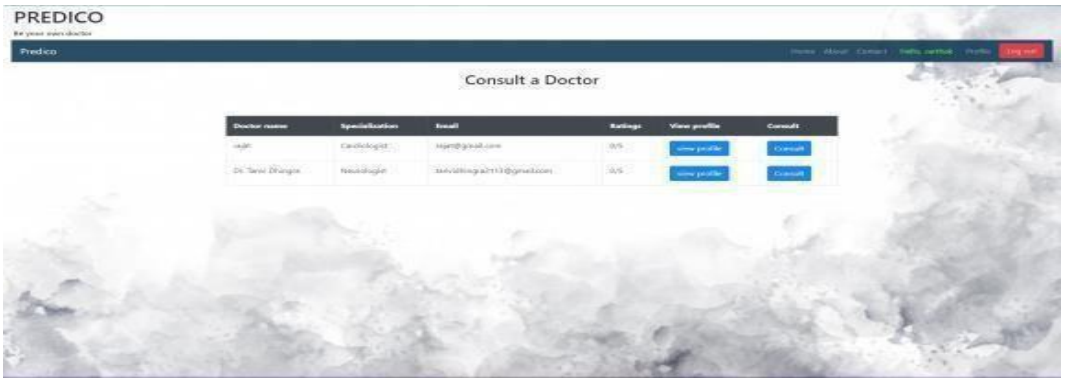


Symptoms Input Page

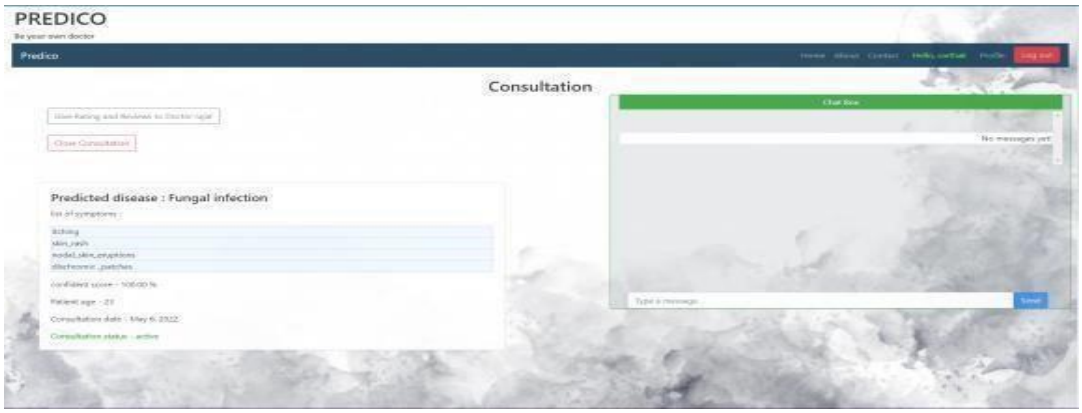


Prediction Result Page

Doctor Consultation Page:



Consultation Chat Page:



ACKNOWLEDGEMENT

Our thanks to the Professors, experts and other faculty members who provided useful resources and background to complete this research paper. The Success and outcome of this project were possible by the guidance and support from many people. We are incredibly privileged to have got this all along with the achievement of this paper. It required a lot of effort from each individual involved in this research paper.

Conclusion

The increasing demand for timely and accurate medical diagnosis has driven the exploration of advanced technologies in the healthcare sector. In this study, we focused on leveraging supervised machine learning algorithms to develop a robust, symptom-based illness prediction system. The central premise was to determine whether these models can significantly enhance the healthcare process by enabling early and precise disease detection. To this end, we investigated a variety of machine learning techniques, including Decision Trees (DT), Random Forest (RF), Gradient Boosting (GB), K-Nearest Neighbors (KNN), and Gaussian Naive Bayes (GNB). Each algorithm was tested and evaluated based on its ability to classify and predict diseases using a set of symptoms provided by users. The results indicate that these models, when properly trained and fine-tuned, can deliver high prediction accuracy, making them viable tools for preliminary diagnosis.

One of the most notable aspects of the proposed system is its interactive design, which allows users to input multiple symptoms and explore additional ones. This feature enhances the model's input quality, leading to more refined disease predictions. By cross-referencing symptoms with a comprehensive database, the system ensures that the output is not only accurate but also relevant to the user's condition.

Furthermore, the study revealed that different algorithms perform better depending on the nature and distribution of the dataset. For instance, ensemble methods like Random Forest and Gradient Boosting often showed superior accuracy and robustness compared to single classifiers. However, simpler models like Gaussian Naive Bayes and KNN also demonstrated competitive results with lower computational costs, making them suitable for lightweight applications or mobile platforms.

In addition to technical accuracy, the user-centered design of the system ensures accessibility and usability, making it a practical solution for individuals seeking a preliminary understanding of their health conditions. This has the potential to reduce the burden on healthcare professionals by filtering non-critical cases and guiding users toward appropriate medical consultations based on their predicted illness.

Overall, the research confirms that supervised machine learning models can play a crucial role in modern healthcare, particularly in disease prediction based on symptoms. While the system is not intended to replace professional medical advice, it can serve as a valuable decision support tool that aids in early detection and encourages timely medical intervention. Future work can focus on expanding the symptom and disease database, integrating realtime patient data, and enhancing the model with deep learning techniques to improve prediction accuracy even further.

REFERENCES:

1) Tikotikar, A., & Kodabagi, M. (n.d.). *A survey on technique for prediction of disease in medical data*. School of Computing & IT, REVA University.

2) Ahmed, Z., Mohamed, K., Zeeshan, S., & Dong, X. (2020). Artificial intelligence with big data analytics for precision medicine in cancer. *Computers in Biology and Medicine, 126*, 104047. https://doi.org/10.1016/j.compbiomed.2020.104047

3) Chen, M., Hao, Y., Hwang, K., Wang, L., & Wang, L. (2017). Disease prediction by machine learning over big data from healthcare communities. *IEEE Access, 5*(1), 8869–8879. https://doi.org/10.1109/ACCESS.2017.2694446

4) Shah, C., & Jivani, A. (n.d.). Comparison of data mining classification algorithms for breast cancer prediction. *IEEE Conference Publication*, IEEE-31661.

5) Balasubramanian, S., & Subramanian, B. (n.d.). Symptom based disease prediction in medical system by using K-means algorithm. *International Journal of Advances in Computer Science and Technology, 3*.

6) Pingale, K., Beyene, C., & Kamat, P. (2019). Disease prediction using machine learning. *International Journal of Pure and Applied Mathematics, 119*(15), 2345–2350.

7) Chitnis, C., & Lee, R. (2018). Improving health-care systems by disease prediction. In *2018 International Conference on Computational Science and Computational Intelligence (CSCI)* (pp. 1344–1349). IEEE. https://doi.org/10.1109/CSCI46756.2018.00260

8) Davis, A., Chawla, D., Blumm, N., Christakis, N., & Barabási, A. L. (2008). Predicting individual disease risk based on medical history.

9) Adam, S., & Parveen, A. (2012). Prediction system for heart disease using Naive Bayes.

10) Groves, P., Kayyali, B., Knott, D., & Kuiken, S. V. (2016). The 'big data' revolution in healthcare: Accelerating value and innovation. *McKinsey & Company*.

11) Rouse, M. (2018). Data mining. *SearchSQLServer*. Retrieved September 12, 2018, from https://searchsqlserver.techtarget.com/definition/data-mining

12) Neelima98. (n.d.). Disease prediction using machine learning [Dataset]. *Kaggle*. Retrieved from https://www.kaggle.com/neelima98/disease-prediction-using-machine-learning

13) Brownlee, J. (2016, March 16). Supervised and unsupervised machine learning algorithms. *Machine Learning Mastery*. Retrieved September 12, 2018, from https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/

14) Scikit-learn. (n.d.). Decision trees. Retrieved September 12, 2018, from https://scikit-learn.org/stable/modules/tree.html

15) Scikitlearn.(n.d.).RandomForestClassifier.RetrievedSeptember12,2018,fromhttps://scikitlearn.org/stable/modules/generated/sklearn.ensemble.Ra ndomForestClassifier.html

16) Scikitlearn.(n.d.).GradientBoostingClassifierRetrievedSeptember12,2018,fromhttps://scikitlearn.org/stable/modules/generated/sklearn.ensemble. GradientBoostingClassifier.html