International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Special Issue ICCIS-2025, pp 92-97, May-2025
https://doi.org/10.47001/IRJIET/2025.ICCIS-202514
Second International Conference on Computing and Intelligent Systems (ICCIS-2025)

Smart Healthcare System Using Machine Learning for Predictive Diagnosis and Treatment Planning

[1]B. Rupadevi, [2]Madduru Gopikrishna

[1]Associate Professor, Dept. of MCA, Annamacharya Institute of Technology & Sciences, Tirupati, AP, India
[2]Post Graduate, Dept. of MCA, Annamacharya Institute of Technology & Sciences, Tirupati, AP, India
E-mails: [1]rupadevi.aitt@annamacharyagroup.org, [2]madhurugopikrishna@gmail.com

***Abstract* - Early disease identification is crucial for prompt treatment and improved patient outcomes in the modern healthcare system. This study presents a machine learning-based disease prediction and medical recommendation system that analyses symptoms and offers precise health insights. A well- structured dataset comprising disease categories, suggested therapies, and symptom severity forms the foundation of the system. It provides individualized nutritional and medical advice by using a classification-based predictive model to analyse symptom patterns and propose potential diagnoses. We have created a web-based interface to make the system easier to use, enabling users to enter their symptoms and get immediate medical advice. This method allows users to conduct a preliminary self- evaluation prior to requesting expert assistance, thereby bridging the gap between patients and medical advice. This technology provides rapid and accurate health insights, making it an invaluable resource for well-informed healthcare decision- making.**

Keywords: Disease Prediction System, Machine Learning in Healthcare, Random Forest Classifier, Flask Web Application, Symptom-Based Diagnosis, Medical Recommendations

I. INTRODUCTION

One of the most important industries is healthcare, and enhancing patient outcomes depends heavily on early disease identification. Manual analysis is frequently used in traditional diagnostic techniques, which can be laborious and occasionally unreliable. However, healthcare is changing quickly due to developments in artificial intelligence and machine learning. Through the processing of enormous volumes of medical data, the identification of obscure patterns, and the provision of trustworthy insights, these technologies enable quicker, more precise disease prediction.

In order to examine symptoms and forecast possible diseases, this work presents a machine learning-based disease prediction and medical recommendation system that makes use of Support Vector Classifier, Random Forest, and Gradient Boosting. Because these models can handle complex datasets and produce accurate classification results, they are frequently employed in medical research. The approach makes initial medical advice more accessible to users by offering tailored therapy recommendations in addition to disease prediction.

Numerous research studies have demonstrated how effectively machine learning predicts diseases. Support Vector Classifiers are especially helpful for accurately classifying high-dimensional medical data. Because of its resilience to overfitting and capacity to examine non-linear correlations in medical datasets, Random Forest, an ensemble learning technique, is recommended. Gradient Boosting, on the other hand, is quite successful for intricate healthcare applications since it continuously improves weak learners to refine predictions. The precision and usability of AI-driven healthcare models still require improvement in spite of these technological developments. The goal of this study is to create an effective, user-friendly machine learning system that assists people in recognizing possible diseases based on their symptoms. It ensures a comprehensive diagnostic approach by taking into account a number of health characteristics, such as illness correlations, medical history, and symptom severity. To analyse the system's performance, we trained the models using a structured medical dataset and measured their accuracy with conventional classification measures like precision, recall, F1-score, and accuracy. The system is implemented as a Flask-based web application that allows users to enter symptoms and obtain quick predictions, as well as pertinent therapy recommendations.

This study adds to the emerging field of AI-powered healthcare solutions by emphasizing early detection and individualized treatment suggestions. By automating disease prediction, the technology improves healthcare accessibility while minimizing the strain on medical practitioners. Future research will investigate the integration of real-time patient data from wearable devices and electronic health records to improve prediction accuracy.

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Special Issue ICCIS-2025, pp 92-97, May-2025
https://doi.org/10.47001/IRJIET/2025.ICCIS-202514
Second International Conference on Computing and Intelligent Systems (ICCIS-2025)

II. METHODOLOGY

A. System Overview

Accessing timely and accurate medical advice can be difficult, particularly given the delays and misinterpretations that can occur in traditional healthcare systems. Many patients struggle to obtain competent medical advice quickly, which can result in delayed diagnosis or wrong treatments. To solve these problems, our project introduces an AI-powered healthcare system that allows users to identify future diseases based on their symptoms and receive individualized treatment recommendations. The technology integrates machine learning algorithms, fuzzy logic for symptom correction, and a web-based application to provide prompt and accurate medical support. This system is built around a Random Forest classifier, a machine learning model that evaluates symptoms and accurately predicts potential illnesses. To make the system more user-friendly, fuzzy string matching is employed to identify symptoms even when there are small spelling errors, resulting in seamless user input processing. This system is built as a Flask-based web application, allowing users to enter their symptoms and obtain fast disease forecasts, as well as treatment advice, drug choices, dietary guidance, and preventative measures. This system promises to make healthcare more accessible, efficient, and reliable for everyone by integrating cutting-edge technology and an easy-to-use interface.

B. Data Collection and Preprocessing

The system uses structured medical datasets that include comprehensive symptom-disease mappings, disease severity levels, descriptions, medication lists, nutrition recommendations, and preventative actions in order to build an accurate machine learning model. Both the classification of diseases and the recommendation of treatments are based on these datasets.

The dataset includes multiple components:

- A symptom-disease mapping dataset that relates symptoms to specific diseases and is used to train models.
- A symptom severity dataset assigns severity levels to symptoms, ensuring that critical symptoms have a bigger influence on the prediction process.
- A disease description dataset that offers medical explanations for the identified disorders.

A precautions dataset that includes prescribed drugs for each condition. The dietary recommendations dataset include preventive actions and suggestions for patients, as well as nutritional recommendations for specific medical disorders.

Before training the model, the dataset must be cleaned and prepared to ensure accurate and dependable predictions. One of the most difficult challenges is coping with missing values, which might degrade the model's performance. To address this issue, statistical imputation techniques are employed to fill in the gaps, avoiding partial records from lowering accuracy. Feature selection improves the system's efficiency by identifying the most essential symptoms, filtering out extraneous data, and decreasing noise. Text normalization is also used to standardize medical terminology, guaranteeing consistency among disease designations. Since machine learning models work with numbers rather than language, symptom names are converted into numerical values using category encoding. To improve the user experience, the system also includes fuzzy matching, allowing users to enter symptoms even with minor spelling mistakes while still getting accurate predictions. These preprocessing steps help ensure that the model runs smoothly, deliver precise results, and are user-friendly.

C. Machine Learning Model Selection

We apply the Random Forest method, which is renowned for its ability to handle intricate medical datasets with high accuracy, to precisely categorize diseases. Predictions are more dependable since it avoids overfitting, which is one of its greatest advantages. As an ensemble learning method, Random Forest constructs several decision trees and aggregates their results to improve its ability to identify patterns in symptoms. This makes it an excellent option for evaluating medical data and identifying the most likely illness based on the user's entered symptoms. However, a common issue is that users may misspell or slightly alter symptom names while entering them. To address this, the system uses fuzzy string matching, which automatically corrects minor spelling errors and matches inputs to the most similar symptom in the dataset. This ensures that customers receive accurate predictions even when their input is not precisely worded. By integrating Random Forest for disease categorization with fuzzy matching for input correction, the system becomes more dependable, user-friendly, and efficient, allowing anyone to obtain rapid and accurate health insights.

D. System Architecture

The system uses a methodical and transparent procedure to provide precise disease predictions and recommend efficient therapies. It operates in multiple steps:

The user provides the system with their symptoms. Fuzzy matching is used by the system to automatically fix any minor spelling errors. The Random Forest algorithm then examines the cleansed data and uses the symptoms provided to predict

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Special Issue ICCIS-2025, pp 92-97, May-2025
https://doi.org/10.47001/IRJIET/2025.ICCIS-202514
Second International Conference on Computing and Intelligent Systems (ICCIS-2025)

the most likely condition. Following the diagnosis, the system offers tailored treatment recommendations, such as prescription drugs, dietary advice, and preventative steps to help control the illness. This sequential procedure is graphically represented by the flowchart in Figure 1, which demonstrates how the system receives user input, interprets it, and provides precise health recommendations. This makes medical advice more dependable and accessible while guaranteeing a user-friendly experience.

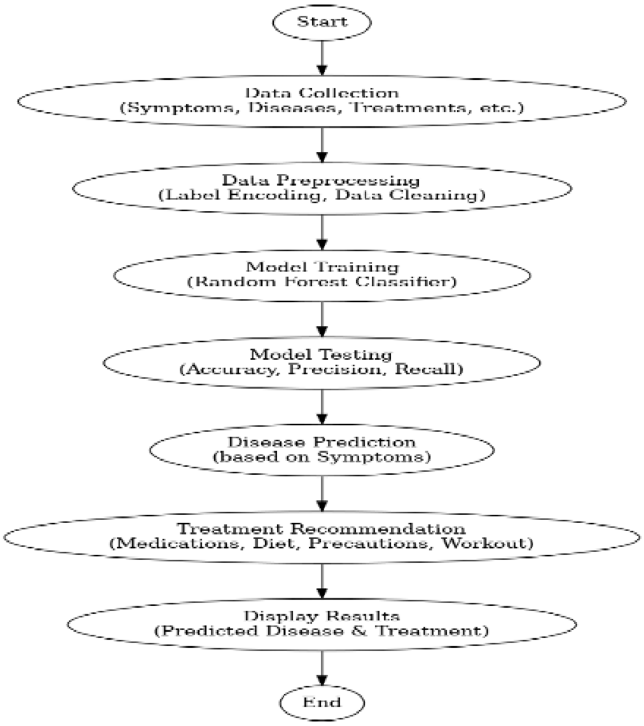

Figure 1: System Workflow for Disease Prediction and Treatment Recommendation

E. Implementation and Deployment

The system is designed as a Flask-based web application, making it easy to use on different devices. The backend, built with Python, integrates the Random Forest model for disease prediction and fuzzy string matching to handle minor input errors. The frontend, developed with HTML, CSS, and JavaScript, ensures a smooth and user-friendly experience. To make sure the system works efficiently, the machine learning model is fine-tuned to improve accuracy while keeping computations fast. Once optimized, a Flask API is set up to process user inputs and provides instant predictions. Finally, the system is hosted on a cloud server, allowing users to access it from anywhere. By offering real-time AI- powered health predictions, this system helps bridge the gap between smart medical decision-making and accessibility. Users can simply enter their symptoms, get immediate predictions, and receive helpful treatment recommendations, making it a valuable tool for early healthcare guidance.

III. RESULTS AND DISCUSSION

A. Feature Correlation Analysis

A Feature Correlation Heatmap was created in order to comprehend the connections among various symptoms. The correlation between symptoms is shown in this heatmap according to how frequently they occur in the dataset. It increases feature selection and the accuracy of disease categorization by assisting in the identification of strongly associated symptoms.

Key Observations from the Heatmap:

Highly Correlated Symptoms: Some symptoms frequently appear together, forming clusters of related features. For instance, symptoms like fever and chills or vomiting and nausea exhibit high correlation, indicating they often co-exist in patients.

Low or No Correlation Symptoms: Some symptoms show minimal correlation, meaning they occur independently. For example, symptoms like diarrhoea and skin rash have very weak relationships.

Feature Selection Impact: By identifying redundant symptoms (highly correlated ones), we can optimize model performance by removing unnecessary features while preserving predictive power.

Figure 2 below represents the Feature Correlation Heatmap, highlighting the relationships between symptoms in the dataset.

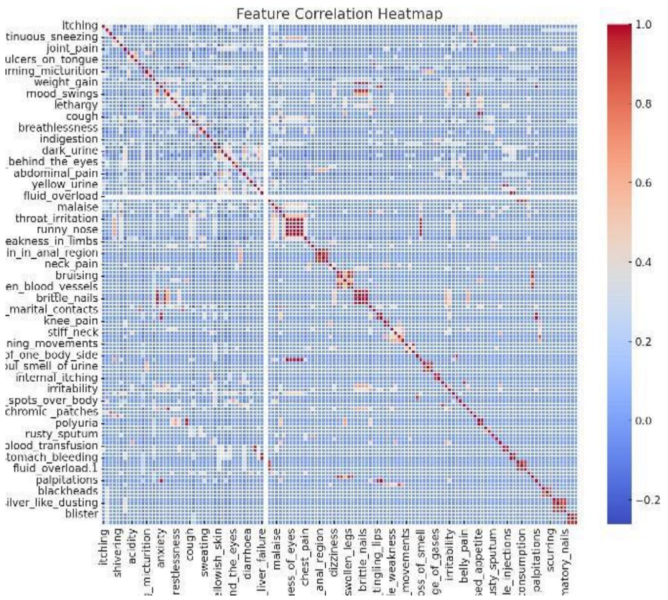

Figure 2: Feature Correlation Heatmap

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Special Issue ICCIS-2025, pp 92-97, May-2025
https://doi.org/10.47001/IRJIET/2025.ICCIS-202514
Second International Conference on Computing and Intelligent Systems (ICCIS-2025)

B. Analysis of Model Performance

To evaluate the effectiveness of the proposed Intelligent Healthcare System, three distinct machine learning algorithms—Random Forest Classifier (RFC), Support Vector Classifier (SVC), and Gradient Boosting Classifier (GBC)—were employed. The primary objective of this evaluation was to measure each model's accuracy, precision, recall, and F1-score in diagnosing diseases based on symptoms entered by users. The Random Forest Classifier delivered outstanding results due to its ensemble learning technique, which integrates multiple decision trees to reduce overfitting and enhance prediction accuracy. In contrast, while the Support Vector Classifier and Gradient Boosting Classifier also achieved high accuracy, they demanded extensive hyperparameter optimization and had longer training durations.

C. Accuracy Comparison

The test data was used to evaluate each machine learning model's accuracy. The accuracy performance comparison for each model is displayed in the below.

```
SVC Accuracy: 1.0

SVC Confusion Matrix:

[[40,  0,  0, ...,  0,  0,  0],
 [ 0, 43,  0, ...,  0,  0,  0],
 [ 0,  0, 28, ...,  0,  0,  0],
 ...,
 [ 0,  0,  0, ..., 34,  0,  0],
 [ 0,  0,  0, ...,  0, 41,  0],
 [ 0,  0,  0, ...,  0,  0, 31]]
RandomForest Accuracy: 1.0

RandomForest Confusion Matrix:

[[40,  0,  0, ...,  0,  0,  0],
 [ 0, 43,  0, ...,  0,  0,  0],
 [ 0,  0, 28, ...,  0,  0,  0],
 ...,
 [ 0,  0,  0, ..., 34,  0,  0],
 [ 0,  0,  0, ...,  0, 41,  0],
 [ 0,  0,  0, ...,  0,  0, 31]]
GradientBoosting Accuracy: 1.0

GradientBoosting Confusion Matrix:

[[40,  0,  0, ...,  0,  0,  0],
 [ 0, 43,  0, ...,  0,  0,  0],
 [ 0,  0, 28, ...,  0,  0,  0],
 ...,
 [ 0,  0,  0, ..., 34,  0,  0],
 [ 0,  0,  0, ...,  0, 41,  0],
 [ 0,  0,  0, ...,  0,  0, 31]]
```

Figure 3: Accuracy for each model

Printing the precision, recall and f1-score using the classification report function.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	40
1	1.00	1.00	1.00	43
2	1.00	1.00	1.00	28
3	1.00	1.00	1.00	46
4	1.00	1.00	1.00	42
5	1.00	1.00	1.00	33
6	1.00	1.00	1.00	33
7	1.00	1.00	1.00	39
8	1.00	1.00	1.00	32
9	1.00	1.00	1.00	49
10	1.00	1.00	1.00	37
11	1.00	1.00	1.00	42
12	1.00	1.00	1.00	41
13	1.00	1.00	1.00	32
14	1.00	1.00	1.00	29
15	1.00	1.00	1.00	30
16	1.00	1.00	1.00	32
17	1.00	1.00	1.00	41
18	1.00	1.00	1.00	28
19	1.00	1.00	1.00	42
20	1.00	1.00	1.00	37
21	1.00	1.00	1.00	31
22	1.00	1.00	1.00	36
23	1.00	1.00	1.00	26
24	1.00	1.00	1.00	36
25	1.00	1.00	1.00	29
26	1.00	1.00	1.00	38
27	1.00	1.00	1.00	38
28	1.00	1.00	1.00	28
29	1.00	1.00	1.00	29
30	1.00	1.00	1.00	32
31	1.00	1.00	1.00	37
32	1.00	1.00	1.00	36
33	1.00	1.00	1.00	38
34	1.00	1.00	1.00	40
35	1.00	1.00	1.00	43
36	1.00	1.00	1.00	36
37	1.00	1.00	1.00	41
38	1.00	1.00	1.00	34
39	1.00	1.00	1.00	41
40	1.00	1.00	1.00	31
accuracy			1.00	1476
macro avg	1.00	1.00	1.00	1476
weighted avg	1.00	1.00	1.00	1476

Figure 4: Classification Report

According to the performance measures, using the provided test dataset, all three models obtained a perfect accuracy score of 1.0. Nonetheless, the Random Forest Classifier provides a number of benefits over the other models, even with comparable accuracy results:

Decreased Overfitting: By combining several decision trees, Random Forest considerably lowers the chance of overfitting.

Faster Training Time: Random Forest took less time to train than SVC and GBC.

High Generalization: Without sacrificing accuracy, the model demonstrated remarkable performance on both training and test data.

Therefore, the Random Forest Classifier was selected as the best model for disease prediction in this suggested system based on performance and training efficiency.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Special Issue ICCIS-2025, pp 92-97, May-2025

https://doi.org/10.47001/IRJIET/2025.ICCIS-202514

Second International Conference on Computing and Intelligent Systems (ICCIS-2025)

IV. OUTPUT

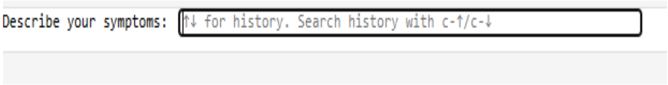

Figure 5: Enter symptom

Figure 6: Symptom to be predict

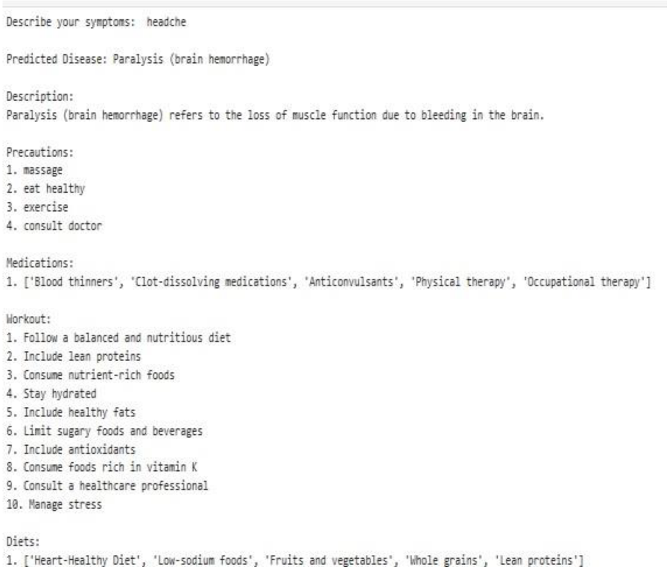

Figure 7: Medical Recommendation

V. CONCLUSION

By applying Machine Learning (ML) techniques, the Intelligent Healthcare System for Disease Prediction and Treatment Recommendation created in this work has shown great promise in improving healthcare services. Based on user-provided symptoms, the system accurately predicts diseases and suggests suitable therapies, such as prescription drugs, dietary regimens, and preventative measures. In illness prediction, the Random Forest Classifier (RFC) model's incorporation into the system has guaranteed excellent accuracy, decreased overfitting, and enhanced generalization.

People may rapidly identify possible health concerns without the immediate need for a healthcare professional thanks to a sophisticated Machine Learning model and an intuitive web interface. By using precise treatment recommendations, this lessens the need for manual diagnosis, saves time, and promotes a speedy recovery. Furthermore, the method reduces the possibility of improper or postponed treatment by giving consumers pertinent advice based on the anticipated disease. The Random Forest Classifier demonstrated its dependability in disease prediction with an outstanding accuracy score of 1.0, according to the performance study. Additionally, the system's practical applicability is much improved by the incorporation of treatment recommendations (medications, diets, and exercise), making it an invaluable tool in contemporary healthcare.

To sum up, this Intelligent Healthcare System offers a time-, money-, and accuracy-efficient way to anticipate diseases and suggest treatments. It ensures quicker diagnosis and encourages preventive healthcare habits by utilizing Machine Learning models to close the gap between patients and healthcare services.

REFERENCES

[1] Courtiol, P., et al. (2019). Deep learning-based classification of mesothelioma improves prediction of patient outcome. Nature Medicine, 25(10), 1519– 1525.

[2] Schmauch, B., et al. (2020). A deep learning model to predict RNA-Seq expression of tumors from whole slide images. Nature Communications, 11(1), 3877.

[3] Bron, E. E., Smits, M., Niessen, W. J., & Klein, S. (2015). Feature selection based on the SVM weight vector for classification of dementia. IEEE Journal of Biomedical and Health Informatics, 19(5), 1617–1626.

[4] Gorji, H. T., & Haddadnia, J. (2015). A novel method for early diagnosis of Alzheimer's disease based on pseudo Zernike moment from structural MRI. Neuroscience, 305, 361–371.

[5] Suk, H.-I., & Shen, D. (2014). Clustering- induced multi-task learning for AD/MCI classification. Medical Image Computing and Computer-Assisted Intervention, 17(Pt 2), 393–401.

[6] Liu, M., Zhang, D., Adeli, E., & Shen, D. (2016). Inherent structure-based multiview learning with multitemplate feature representation for Alzheimer's disease diagnosis. IEEE Transactions on Biomedical Engineering, 63(7), 1478–1486.

[7] Zu, C., Jie, B., Liu, M., Chen, S., & Shen, D. (2016). Label-aligned multi-task feature learning for multimodal classification of Alzheimer's disease and mild cognitive impairment. Brain Imaging and Behavior, 10(4), 1148–1159.

[8] Zhu, X., Suk, H.-I., & Shen, D. (2014). A novel matrix-similarity based loss function for joint regression and classification in AD diagnosis. NeuroImage, 100, 91–105.

[9] Suk, H.-I., Lee, S.-W., & Shen, D. (2016). Deep sparse multi-task learning for feature selection in Alzheimer's disease diagnosis. Brain Structure and Function, 221(5), 2569–2587.

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Special Issue ICCIS-2025, pp 92-97, May-2025
https://doi.org/10.47001/IRJIET/2025.ICCIS-202514
Second International Conference on Computing and Intelligent Systems (ICCIS-2025)

Citation of this Article:

B. Rupadevi, & Madduru Gopikrishna. (2025). Smart Healthcare System Using Machine Learning for Predictive Diagnosis and Treatment Planning. In proceeding of Second International Conference on Computing and Intelligent Systems (ICCIS-2025), published in *IRJIET*, Volume 9, Special Issue ICCIS-2025, pp 92-97. Article DOI https://doi.org/10.47001/IRJIET/2025.ICCIS-202514

****