International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 14, Issue 2, 2026

AI Health Symptoms Checker & Disease Predictor

Mrs B. Hanumantha Rao[1], Kola Praneetha , Pisini Sri Kavya[2], Syed Sha Karimulla[3], Gunti Venkata Ratnam[4]

[1] Assistant Professor of CSE,Computer Science Engineering Department,Potti Sriramulu Chalavadi Mallikarjuna Rao College Of Engineering,One Town ,Vijayawada, India.

[2,3,4,5]Students;Computer Science Engineering Department,Potti Sriramulu Chalavadi Mallikarjuna Rao College Of Engineering, One Town ,Vijayawada, India

Mail Id; bhr@pscmr.ac.in(guide), 22kt1a0514@pscmr.ac.in, kavyapisini7@gmail.com, syedshakarimulla111@gmail.com, venkatarathnamgunti@gmail.com.

Abstract: The rapid growth of artificial intelligence in healthcare has created new opportunities for early disease detection and decision support systems. This paper presents an AI-Driven Health Symptom Checker and Intelligent Disease Prediction System that predicts possible diseases based on symptoms entered by users. The proposed system aims to provide quick preliminary health assessment, especially for users with limited access to medical professionals. A structured medical dataset containing symptoms and corresponding diseases is used to train the prediction models.

The system applies multiple machine learning algorithms such as Decision Tree, Random Forest, Gradient Boosting, and Naive Bayes to analyze symptom patterns and improve prediction accuracy. Data preprocessing, feature encoding, model training, and performance evaluation are performed to identify the most effective model. A user-friendly web interface is integrated to allow users to input symptoms and receive instant predictions in real time. Experimental results show that the proposed system provides reliable and efficient disease prediction with reduced response time. The system can support early diagnosis, improve health awareness, and act as a helpful assistant before professional consultation.

Index terms **-** *Artificial Intelligence, Disease Prediction, Symptom Checker, Machine Learning, Healthcare Analytics, Random Forest, Decision Tree, Naive Bayes, Gradient Boosting, Web Application, Early Diagnosis, Predictive Healthcare*

1. INTRODUCTION

Artificial Intelligence (AI) has become one of the most transformative technologies in the healthcare sector, offering intelligent solutions for diagnosis, prediction, and decision-making. With the increasing population and growing demand for healthcare services, many individuals face difficulty in obtaining timely medical consultation. Early identification of diseases plays a crucial role in effective treatment, but lack of awareness, delayed diagnosis, and limited access to healthcare professionals often lead to serious complications. Therefore, there is a strong need for automated systems that can assist users in understanding their health conditions at an early stage.

The proposed AI-Driven Health Symptom Checker and Intelligent Disease Prediction System is developed to predict possible diseases based on symptoms entered by the user. The system uses machine learning techniques to analyze symptom patterns and provide accurate predictions in real time. A structured dataset containing symptoms and related diseases is used for training multiple classification models such as Decision Tree, Random Forest, Gradient Boosting, and Naive Bayes. These algorithms help in identifying complex relationships between symptoms and diseases more effectively than traditional rule-based systems.

A user-friendly web interface is integrated into the system, allowing users to enter symptoms easily and receive instant health predictions. The trained models are stored and reused for efficient real-time performance. This system can act as a preliminary healthcare assistant by improving disease awareness and encouraging timely medical consultation. Hence, the project demonstrates how AI can enhance healthcare accessibility, reduce diagnosis delays, and support smart medical decision-making.

2. LITERATURE SURVEY

1. AI Doctor Symptom Checker with Recommendation

The necessity for systems that can evaluate patient symptoms in real time and offer trustworthy advice is highlighted by the growing need for sophisticated and easily available healthcare solutions. The creation of AI Doctor—Symptom Checker with Recommendation, a web-based platform driven by machine learning for early symptom analysis and health guidance, is presented in this article. The program incorporates AI-powered prediction algorithms, secure authentication, and customized dashboards. The system is built with HTML/CSS/JavaScript for the frontend and Python (Flask) for the backend, guaranteeing accuracy, responsiveness, and usability. The outcome validates

accurate illness prediction, easy-to-use interface, and safe data processing. The solution delivers organized suggestions, increased interaction, and the possibility of AI-driven customisation as compared to conventional symptom checks.

2. Perceived Challenges to Effective Pediatric Palliative Care in India: A Systematic Review

Background

Pediatric palliative care improves life-threatening illness patients' quality of life. Pediatric palliative care rules and methods differ substantially from adult ones. Indian palliative care is growing, however pediatric palliative care is still in its infancy. This systematic study examined India's pediatric palliative care difficulties.

Method

The articles were picked using various search phrases and parameters. Three databases yielded 482 entries. 10 papers were reviewed from 55 selected for full-text analysis. PRISMA was used to pick articles.

Results

The current evidence shows the need for pediatric palliative care nationwide. Challenges include a lack of training, knowledge, NGOs providing care, especially for children, stigma, resources, and child-specific policies.

Conclusion

Multiple problems in pediatric palliative care suggest its successful health care system implementation. These problems can improve pediatric palliative care for children and their families if managed well.

3. Disease Prediction Using Machine Learning

Electronic data accumulated as a result of the health care sector's widespread adoption of computer-based technology. Medical professionals are having difficulty effectively analyzing symptoms and detecting illnesses early on because of the large volumes of data. Supervised machine learning (ML) algorithms, on the other hand, have demonstrated considerable promise in outperforming conventional disease diagnosis systems and supporting medical professionals in the early identification of high-risk disorders. The goal of this research is to identify patterns in illness diagnosis across different kinds of supervised machine learning models by analyzing performance indicators. The most widely discussed supervised machine learning algorithms were K-Nearest Neighbor (KNN), Decision Trees (DT), and Naïve Bayes (NB). According to research, Support Vector Machine (SVM) is the best method for identifying Parkinson's disease and renal disorders. When it came to predicting cardiac conditions, Logistic Regression (LR) did quite well. Lastly, precision breast illnesses were predicted by Random Forest (RF) and common diseases by Convolutional Neural Networks (CNN).

4. Symptom-Based Disease Prediction: A Machine Learning Approach

The development of machine learning methods has transformed a number of industries, including healthcare. The focus of this study is using machine learning algorithms to forecast diseases based on symptoms. The goal is to create a reliable predictive model that can correctly identify diseases based on a collection of symptoms using a dataset that includes 132 symptoms and 41 diseases. There are various crucial phases in the procedure. To manage missing values, encode category variables, and standardize the data, the dataset is first preprocessed. Feature selection approaches are used to identify which symptoms are most relevant to a disease's prognosis. To find the best prediction model, a number of machine learning methods were investigated, such as decision trees, support vector machines, random forests, and XGBoost. Because of its ability to handle complex interactions within the data and its efficacy in managing unbalanced datasets, XGBoost in particular stands out as one of the top-performing models. Evaluation measures including as accuracy, precision, recall, and F1-score are used to assess the models' performance. Additionally, methods like cross-validation and hyperparameter tweaking are used to improve model performance and prevent overfitting. The suggested method has the potential to significantly improve patient outcomes and save healthcare costs by helping medical practitioners diagnose illnesses quickly and correctly. It is crucial to remember that in order for the model to continue to be useful in clinical settings, it requires additional validation on a variety of datasets and frequent updates.

5. Improving the accuracy of medical diagnosis with causal machine learning

Clinical diagnosis and decision-making might be completely transformed by machine learning. The goal of a medical diagnostic is to identify the illnesses that are producing a patient's symptoms. Nevertheless, current machine learning methods for diagnosis are just associative, detecting illnesses that have a high correlation with a patient's symptoms. We demonstrate how this incapacity to separate correlation from causality might lead to harmful or less-than-ideal diagnoses. In order to get around this, we construct counterfactual diagnostic algorithms and reformulate diagnosis as a counterfactual inference job. Using a test set of clinical vignettes, we evaluate our counterfactual algorithms against 44 physicians and the traditional associative method. Our counterfactual approach achieves expert clinical accuracy in the top 25% of physicians, whereas the associative method places in the top 48% of physicians in our dataset. Our findings demonstrate

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 14, Issue 2, 2026

that causal reasoning is a crucial component that is lacking when using machine learning for medical diagnosis.

3. METHODOLOGY

i) Proposed Work:

The proposed work focuses on developing an AI-Driven Health Symptom Checker and Intelligent Disease Prediction System that can predict possible diseases based on symptoms entered by the user. The system is designed to overcome the limitations of traditional diagnosis methods by providing quick, accurate, and real-time predictions through a web-based platform. A structured dataset containing various symptoms and corresponding diseases is used to train the system for intelligent healthcare assistance.

The proposed model applies multiple machine learning algorithms such as Decision Tree, Random Forest, Gradient Boosting, and Naive Bayes to analyze symptom patterns and compare prediction performance. After preprocessing and encoding the dataset, the best-performing model is selected and stored using joblib for future use. Users can input symptoms through an interactive interface, and the trained model instantly predicts the most probable disease. This system helps in early diagnosis, improves healthcare accessibility, and supports users in taking timely medical decisions.

ii) System Architecture:

The system architecture of the proposed AI-based disease prediction system is designed as a modular framework that ensures smooth data flow from user input to final prediction output. The architecture begins with a web-based user interface, where users can register, log in, and enter their symptoms through an interactive dashboard. The entered symptoms are collected by the input processing module, which validates the data and converts it into a structured machine-readable format suitable for prediction.

The processed input is forwarded to the prediction engine, where pre-trained machine learning models such as Decision Tree, Random Forest, Gradient Boosting, and Naive Bayes are loaded from stored joblib files. The system analyzes symptom combinations and predicts the most probable disease with high accuracy. The predicted result is then sent to the output module and displayed on the user dashboard in real time. The architecture also includes a training module connected to the dataset for model building, testing, and updating, making the system scalable, efficient, and suitable for intelligent healthcare applications.

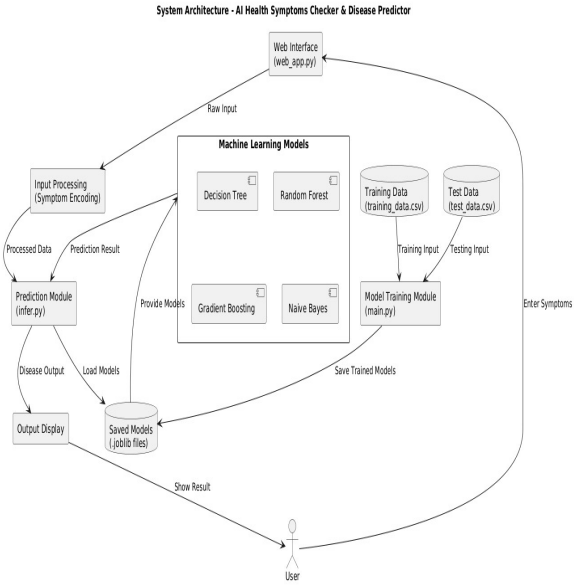

Fig.1. Proposed Architecture

iii) MODULES:

1. User Authentication Module

This module manages user registration, login, and account access. It ensures secure authentication for both admin and users before accessing the system features.

2. Symptom Input Module

Users can select or enter symptoms through the web interface. The module collects user symptoms in an easy and structured format for further processing.

3. Data Preprocessing Module

This module cleans and converts symptom data into numerical encoded format. It ensures the input matches the trained machine learning model requirements.

4. Model Training Module

The system trains multiple machine learning algorithms such as Decision Tree, Random Forest, Gradient Boosting, and Naive Bayes using the healthcare dataset. It helps in selecting the best-performing model.

5. Disease Prediction Module

This is the core module of the system. It receives processed symptoms and predicts the most probable disease using the trained model in real time.

6. Result Display Module

The predicted disease output is shown clearly on the user dashboard. It provides quick results for preliminary health analysis.

7. Model Storage Module

The trained model is saved using joblib files for reuse. This avoids retraining every time and improves prediction speed.

8. Admin Management Module

Admin can monitor users, datasets, and system performance. It helps in maintaining the application efficiently.

iv) ALGORITHMS:

1. Decision Tree Algorithm

Decision Tree is a supervised machine learning algorithm used for classification tasks. It works by dividing the dataset into branches based on symptom conditions and creates a tree-like structure of decisions. Each internal node represents a symptom, each branch represents an outcome, and each leaf node gives the predicted disease. It is easy to understand, requires less preprocessing, and provides fast predictions for healthcare applications.

2. Random Forest Algorithm

Random Forest is an ensemble learning algorithm that combines multiple Decision Trees to improve prediction performance. It trains several trees using random subsets of data and features, then gives the final result through majority voting. This method increases accuracy, reduces overfitting, and handles large datasets efficiently. It is highly suitable for disease prediction where symptoms may overlap across diseases.

3. Gradient Boosting Algorithm

Gradient Boosting is a boosting technique that builds models sequentially, where each new model focuses on correcting the errors made by previous models. It combines multiple weak learners to form a strong predictive model. This algorithm provides high accuracy and can capture complex symptom-disease relationships. It is effective for medical datasets where precise classification is required.

4. Naive Bayes Algorithm

Naive Bayes is a probabilistic machine learning algorithm based on Bayes theorem. It assumes that all symptoms are independent of each other and calculates the probability of a disease based on the given symptoms. It is simple, fast, and efficient for classification tasks. Naive Bayes performs well on categorical and binary symptom datasets used in healthcare prediction systems.

5. Label Encoding Algorithm

Label Encoding is a preprocessing technique used to convert categorical values such as disease names into numerical labels. Since machine learning algorithms require numerical input, this method helps transform textual classes into machine-readable format. It improves compatibility between the dataset and the prediction models.

6. Train-Test Split Technique

Train-Test Split is a data preparation method used to divide the dataset into training and testing sets. The training set is used to train the machine learning models, while the testing set is used to evaluate their performance on unseen data. This technique helps measure accuracy, reliability, and generalization ability of the prediction system.

4. EXPERIMENTAL RESULTS

The proposed AI-Driven Health Symptom Checker and Intelligent Disease Prediction System was implemented and tested using a structured medical dataset containing symptoms and corresponding diseases. The dataset was divided into training and testing sets to evaluate the performance of multiple machine learning algorithms such as Decision Tree, Random Forest, Gradient Boosting, and Naive Bayes. Performance metrics including accuracy, precision, recall, and F1-score were considered to identify the most efficient model. The experimental analysis showed that ensemble-based models such as Random Forest and Gradient Boosting achieved better accuracy compared to single classifiers.

The web-based application was also tested for usability, response time, and real-time prediction capability. Users were able to enter symptoms through the interface and receive instant disease predictions with minimal delay. The stored trained models using joblib reduced computation time and improved system efficiency. Experimental results confirmed that the proposed system provides reliable disease prediction, faster response, and better accessibility, making it suitable for preliminary healthcare assistance and early diagnosis support.

Accuracy: The ability of a test to differentiate between healthy and sick instances is a measure of its accuracy. Find the proportion of analysed cases with true positives and true negatives to get a sense of the test's accuracy. Based on the calculations:

Accuracy = TP + TN /(TP + TN + FP + FN)

$$Accuracy = \frac{(TN + TP)}{T}$$

Precision: The accuracy rate of a classification or number of positive cases is known as precision. Accuracy is determined by applying the following formula:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\Pr e\,cision = \frac{TP}{(TP + FP)}$$

Recall: The recall of a model is a measure of its capacity to identify all occurrences of a relevant machine learning class. A model's ability to detect class instances is shown by the ratio of correctly predicted positive observations to the total number of positives.

$$Recall = \frac{TP}{(FN + TP)}$$

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 14, Issue 2, 2026

F1-Score: A high F1 score indicates that a machine learning model is accurate. Improving model accuracy by integrating recall and precision. How often a model gets a dataset prediction right is measured by the accuracy statistic..

$$F1 = 2 \cdot \frac{(Recall \cdot \Pr e\,cision)}{(Recall + \Pr e\,cision)}$$

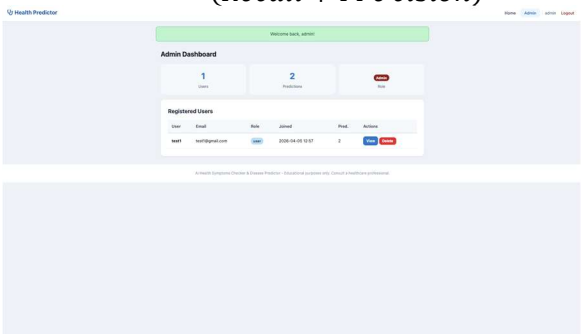

Fig.2. admin dashboard

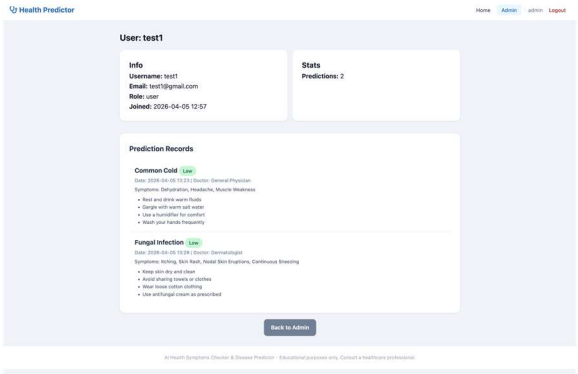

Fig.3. User dashboard

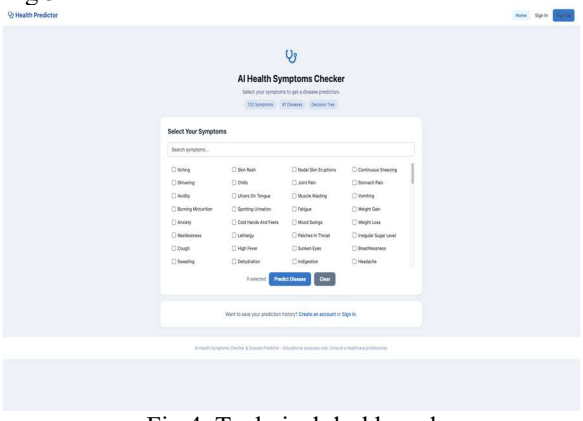

Fig.4. Technical dashboard

5. CONCLUSION

The proposed AI-Driven Health Symptom Checker and Intelligent Disease Prediction System successfully demonstrates the application of artificial intelligence and machine learning in the healthcare domain. By analyzing user-entered symptoms, the system predicts possible diseases accurately and provides quick preliminary health insights. The use of multiple machine learning algorithms such as Decision Tree, Random Forest, Gradient Boosting, and Naive Bayes improved prediction performance and helped identify the most reliable model for real-time deployment.

The developed web-based platform offers a simple and user-friendly interface, enabling users to access disease prediction services easily from anywhere. The system reduces diagnosis delays, increases health awareness, and supports early medical consultation. Overall, the project provides an efficient, scalable, and practical solution that can assist users in making timely health decisions while complementing professional healthcare services.

6. FUTURE SCOPE

The proposed system can be further enhanced by integrating larger and real-time medical datasets to improve prediction accuracy and support a wider range of diseases. Advanced deep learning techniques such as Artificial Neural Networks (ANN), CNN, and hybrid ensemble models can be implemented to analyze complex symptom patterns more effectively. The system can also be upgraded to provide confidence scores, personalized health suggestions, and multilingual support for better user experience.

In future, the application can be deployed as a mobile app and integrated with wearable health devices for continuous monitoring of patient conditions. It can also be connected with hospitals, doctors, and telemedicine platforms for direct consultation after prediction. These enhancements would transform the system into a complete smart healthcare assistant for real-time diagnosis and preventive care.

REFERENCES

1. Harini, V., et al. (2025). AI Doctor Symptom Checker with Recommendation. International Journal of Advanced Research and Review.
2. Sudhakar, K., & Thummuru, P. S. (2025). Disease Prediction System Using Machine Learning Models. International Journal of Engineering Research in Science and Technology (IJERST).
3. Mishra, P., et al. (2025). Disease Prediction Using Machine Learning: A Comparative Study. International Journal for Research in Applied Science and Engineering Technology (IJRASET).
4. Sirigineedi, M., et al. (2024). Symptom-Based Disease Prediction: A Machine Learning Approach. Journal of Artificial Intelligence and Machine Learning.

ISSN 2347–3657
Volume 14, Issue 2, 2026

5. Richens, J. G., Lee, C. M., & Johri, S. (2020). Improving the Accuracy of Medical Diagnosis with Machine Learning. Nature Communications.
6. Reddy, M. V. K., et al. (2022). Disease Predictor Based on Symptoms Using Machine Learning. IJRASET.
7. Sood, R., & Sharma, V. (2025). Symptom Based Disease Prediction Using Machine Learning. International Journal of Preventive Medicine and Health.
8. Scikit-learn Documentation. (2024). Machine Learning in Python. Available at: https://scikit-learn.org
9. Pandas Documentation. (2024). Data Analysis and Manipulation Tool. Available at: https://pandas.pydata.org
10. NumPy Documentation. (2024). Numerical Computing Library. Available at: https://numpy.org
11. Flask Documentation. (2024). Web Development Framework for Python. Available at: https://flask.palletsprojects.com