© SEP 2025 | IRE Journals | Volume 9 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I3-1710625-352

Symptom Based Disease Prediction

PAVAN KUMAR P M[1], DR. KUMAR SIDDAMALLAPPA U[2], SOWMYA P[3]

[1]Student, DOS in Computer Applications (MCA), Davangere University, Shivagangothri, Davanagere, Karnataka, India

[2]Assistant Professor, DOS in Computer Applications (MCA), Davangere University, Shivagangothri, Davanagere, Karnataka, India

[3]Research Scholar, DOS in Computer Science, Davangere University, Shivagangothri, Davanagere, Karnataka, India

Abstract- Predictive disease is an important part of preventive medicine because it reports possible diseases in their immature form just before their symptoms appear. Clinical expertise is frequently involved in the process of manual diagnosis, and is not always available or accurate in a time sensitive scenario. The project, Disease Prediction from Symptoms, is an initiative that uses the practices of Machine Learning to help with determining likely diseases based on the input of symptoms provided by a user to the system. The system is developed in Python, and it has a Tkinter-based GUI to allow a secure login and registration, and an interactive dashboard that allows one to add symptoms. The input is processed by a trained classification model which is trained by applying algorithms like the Random Forest and Decision Tree among others and the most likely disease is estimated. The model has been trained based on a structured data set of wide spectrum of diseases and combination of symptoms. The application does not only prognose the disease but also offers precautionary options and some general medical care tips, which are helpful in early diagnosis and intervention prior to a visit to a health care specialist. This project adds to the accessible, affordable, and reliable health help by offering the combination of information-based learning and user-friendly interaction. The implementation can be improved in future by incorporating electronic health records, real-time monitoring of patient status, and deep learning models to achieve better accuracy.

Index Terms- Disease Prediction, Machine Learning, Symptom Analysis, Healthcare Application

I. INTRODUCTION

In the modern world where time is the key, healthcare systems struggle to deliver efficient diagnosis in time because of the rising number of patients, the lack of medical specialists, and the resemblance of symptoms between some diseases. The importance of early diagnosis of possible diseases is immense since it enables patients to consult doctors before the situation gets out of hand. Nonetheless, manual diagnosis is usually tedious, subjective and can be easily compromised. As the technologies of Machine Learning (ML) and Artificial Intelligence (AI) evolve, automated disease prediction systems have become a promising tool to help medical workers and guide patients with credible advice.

The objective of this project, Disease Prediction from Symptoms, is to come up with an intelligent system that makes the most likely prediction of the disease based on the symptoms that users input into the system. The system learns to behave in the complex patterns within the medical data by training machine learning models, including the Decision Trees, Rand Forests, and Support Vector Machines, and enables the system to make correct predictions. The system receives input in the form of a graphical interface based on Tkinter in which a user can input all the symptoms, after which the predicted disease and precautionary measures are displayed to them.

Such a system is beneficial both to the patient and the doctor since it acts as a tool to aid in decision making. It makes access more open, decreases reliance on urgent medical care, and helps to preventative medicine. It is possible to enhance the predictability and breadth of the disease prediction systems in future by incorporating more data, more sophisticated deep learning strategies and real-time tracking.

II. LITERATURE SURVEY

Fazli et al. [1] This review synthesizes ML approaches used across many disease domains, providing a bibliometric analysis and comparing common algorithms (tree-based models, SVM, neural nets) and data challenges such as class imbalance and explainability. It emphasizes evaluation standards and the need for robust preprocessing when working with heterogeneous clinical inputs — directly relevant when selecting models and validation strategies for symptom-based prediction.

Girma et al. [2] This study builds classifiers to predict disease categories directly from symptom text in a low-resource language, using feature engineering and standard classifiers plus thorough cross-validation. It highlights how careful text preprocessing and class-balancing improve multiclass performance — useful if your app accepts free-text or non-standard symptom descriptions.

Ahsan, M. M et al. [3] This paper explores NLP pipelines (tokenization, TF-IDF/embeddings) combined with deep models to map symptom descriptions to disease labels. The authors report that embedding + neural classifiers outperform simple bag-of-words for noisy, user-entered symptom descriptions — an approach to consider if you plan to accept natural language symptom input.

Amisha Tirkey et al.[4] The authors compare SVM, Random Forest, and Naive Bayes on a curated symptom–disease dataset and report competitive accuracies (often >85–90%) after targeted feature selection and data cleaning. They stress building an extensible symptom vocabulary and GUI integration for real-world usability — directly aligning with a Tkinter front end and structured symptom lists.

Nature Medicine / Few-shot phenotyping work addresses the long-tail problem (rare diseases) by applying few-shot learning to phenotype descriptions, showing that meta-learning helps generalize from very few labeled cases. For your project, the implication is to consider few-shot or transfer techniques if you want to expand coverage to rare conditions with limited examples.

Saha et al.[5] A disease-specific example where extensive feature engineering and comparative evaluation (DT, RF, SVM, NN) on standard cardiac datasets led to improved early-warning performance. This illustrates best practices (feature selection, cross-validation, class-imbalance handling) you can apply when modeling symptom→disease mappings.

Swathi M et al. [6] This recent scoping review surveys ML use across disease prediction tasks, calling out common pitfalls (overfitting on small datasets, lack of external validation) and recommending transparent reporting (confusion matrices, per-class metrics). Use these recommendations when reporting performance for your multi-class symptom-prediction model.

III. PROPOSED METHOD

The proposed system will recommend the most likely disease according to the symptoms given by the user through a machine learning algorithm. This is started by data collection and pre-processing where a structured dataset of various diseases and their related symptoms is cleansed, normalised, and into a format readable by machines. The existence of each symptom is coded as a binary feature (present/absent) and diseases are modeled by target labels.

Once pre-processed, the data is divided into training and testing data. Various machine learning classifiers used include Decision Tree, random Forest and Support Vector Machine (SVM) which are trained to find relations among symptoms and illnesses. Among them, the Random Forest model should offer a better precision, since it engages in ensemble learning and can address large features sets.

A GUI is developed in Tkinter, and it is supposed to enable human interaction, by means of which a person can safely log in and enter symptoms to be shown the predicted diseases and precautionary measures. The trained model is then connected to the GUI, which allows real-time input processing by using the trained model and provides the correct results.

This process provides an effective, scaled and user-friendly disease prediction method. The system helps patients to take preventative steps prior to visiting medical practitioners because of giving timely insights. The extensions planned in the future could involve the use of deep learning models, cloud storage, and real time health monitoring to provide more strong predictions.

3.1 Problem Statement

Limited resources, emerging patient pressures, and overlapping symptoms in various diseases have led to difficulties in healthcare systems all over the world to deliver suitable and precise disease diagnosis on time. There is a tendency of patients not recognizing the gravity of their symptoms and delay seeking medical attention and this may cause problems. Conventional diagnosis is also dependent on the availability of experts, and this might not be in all cases, not mentioning rural or underserved locations.

The issue with machine learning in medical prediction is that despite the promise delivered, there is still no gap in creating user-friendly, accurate, and accessible tools capable of analyzing various symptoms and offering potential disease predictions. Most of the systems in place are not integrated with a simplified interface and thus not user friendly to non-technical users.

Hence, an intelligent, automated system on how user-provided symptoms are processed by machine learning algorithms and how they can be relied on to give the correct prediction of a disease and their precautions is necessary. This solution would strengthen preventive healthcare, dependency on the availability of experts immediately, and help in early decisions made by patients.

3.2 Objectives

The major goals of this project are:

a. To engineer and create a smart system that would predict diseases through machine learning methods by offering information about the diseases through the methods posted by the user.
b. To preprocess and train models like Decision Tree, Rand Forest and SVM on an organized set of symptoms and diseases to successfully categorize them.
c. To create a user-friendly Tkinter-based GUI, which enables secure login, symptom entry and real-time disease prediction.
d. To offer precautionary measures and guidance and predictions, help the users in taking precautionary steps before seeing a medical practitioner.
e. To make healthcare more available to patients in rural or underserved regions where expert medical consultation is not necessarily readily accessible by providing them with a low-cost, user-friendly aid.
f. To compare and assess the performance of the models based on such measures as accuracy, precision, and recall, make sure that the system is reliable and can be scaled.
g. To provide the future extensions as the integration of deep-learning, cloud-based storage, and real-time monitoring of patient health to enhance diagnosis.

3.3 Existing Methods

Conventional healthcare diagnosis is highly dependent on manual assessment of the medical practitioners, in which the analysis of symptoms depends on the patient history, physical examination, and clinical experience. Though effective, this method is time consuming, subjective and largely relies on the access and competence of health care personnel. In most of the rural or underserved areas, there is a lack of access to medical professionals in time thus hindering timely diagnosis and treatment.

Simple decision-support tools and rule-based expert systems have been created in recent years to assist in diagnosis. Such systems are typically based on preset symptom disease mappings or decision rules. They however, do not usually deal with overlapping symptoms, multi relationships or larger scale changes in data.

There are some models of machine learning already in existence, such as Naive Bayes, Decision Trees and Support Vector Machines to use in disease prediction using symptoms. These models are moderately accurate, but also have not user-friendly interfaces and have not been developed to be used in real-time, interactively. In addition, most of these current systems lack precautionary measures or recommendations and hence cannot be practically used by patients.

Hence, the current systems offer partial solutions, but do not offer scalability, accuracy or accessibility. This introduces a loophole which is filled by the proposed system as it combines advanced machine learning methods with a user friendly Tkinter based GUI, providing both predictive and precautionary advice on-the-fly.

3.4 Implementation

The Disease Prediction from Symptoms system implementation is a multi-task process, starting with data preprocessing all the way to the model execution with an easy interface.

Dataset Preparation

- A tabular set of symptoms and diseases were gathered.
- Target labels were diseases each coded as binary (1 present and 0 absent), and each symptom was coded as binary (1 present and 0 absent).
- The data was cleaned to eliminate the inconsistency and missing values.

Data Preprocessing

- Normalized features (symptoms) were made.
- The data were divided into training and testing data sets.
- The dimensionality reduction methods were used in order to eliminate redundant features.

Model Training

- Decision Tree, Random Forest, and Support Vector Machine (SVM) algorithms of machine learning were used.
- The training dataset was used to train each of the models and test on the test dataset.
- The Raeforest model was the most accurate because of the ability of ensemble learning and capacity to run large sets of features.

Graphical User Interface (GUI).

- A GUI based on Tkinter was created in order to give an interactive platform.
- There are features; login and registration, a dashboard and a symptom entry form.
- The system recommends the most likely disease and shows precautionary actions when the user enters the symptoms.

Integration & Testing

- Trained model was installed with the GUI to make real time predictions.
- To achieve reliability and accuracy in the system, the system was evaluated using varied combinations of symptoms.

3.5 Snapshot

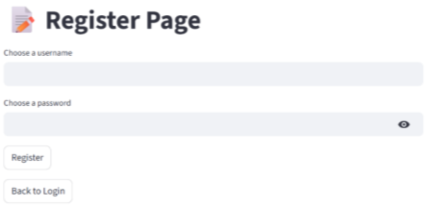


Fig.1: Register Page

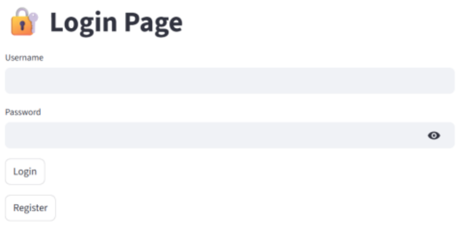
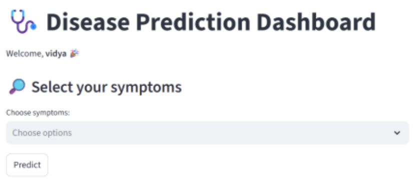


Fig.2: Login page



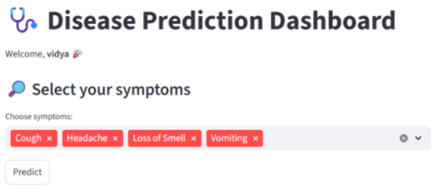
Disease Prediction Dashboard

Welcome, vidya

Select your symptoms

Choose symptoms: Choose options

Fig 3: Dashboard



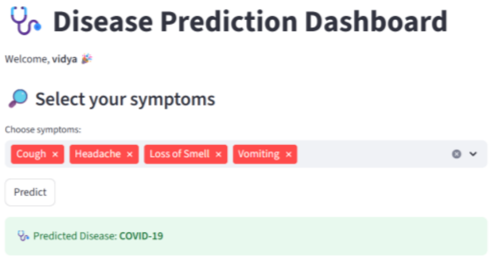
Disease Prediction Dashboard

Welcome, vidya

Select your symptoms

Choose symptoms: Cough × Headache × Loss of Smell × Vomiting ×

Fig..4: Symptoms Selection



Disease Prediction Dashboard

Welcome, vidya

Select your symptoms

Choose symptoms: Cough × Headache × Loss of Smell × Vomiting ×

Predicted Disease: COVID-19

Fig.5: Predicted Disease

CONCLUSION

Disease Prediction from Symptoms shows that machine learning has the capacity to improve prophylactic medical care and help with early diagnosis. With properly trained models like Decision Tree, Random Forest and SVM, the system can make an accurate prediction on the most likely disease with a reasonable level of certainty, using a structured symptom-disease dataset. The Random Forest model was the best amongst these as it proved to be more robust and also because of its assembly learning mechanism. Its interactive and user-friendly interface due to the integration of a Tkinter-based GUI enables the system to enable people to securely log in and input their symptoms and obtain not only the prediction of the disease but also precautionary measures. This fills the gap between cutting edge machine learning technologies and real-world healthcare support, particularly in rural and underserved areas where access to medical expertise cannot be available immediately.

Although the existing implementation fulfils the main goal of prediction of the disease, it also provides the opportunity to enhance it with deep learning, free-text analysis of symptoms, deployment on the cloud, and multimodal health monitoring. On the whole, this system emphasizes the potential artificial intelligence use in healthcare to offer affordable, scalable, and efficient decision-support solutions and eventually lead to improved health awareness and early medical intervention.

FUTURE WORK

Although the presented system proves to be effective in terms of predicting the disease through the symptoms, there still remains much room of improvement and growth. Improvements that can be made in the future are:

- Deep Learning Model Integration - The introduction of the advanced architecture of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to detect the more intricate patterns and enhance the precision of the prediction.
- Greater and Real-World Datasets - Scale up the dataset with actual patient data, disparate demographics and uncommon illnesses to raise dependability and broad applicability.
- Natural Language Processing (NLP) – NLP enables users to enter symptoms in free-text mode, which can be run through NLP pipelines to provide more flexible interactions, which are more realistic.
- Cloud and Mobile Integration - Implementing the system on cloud and mobile systems in order to enhance accessibility and real-time availability of more users.

- Personalized Recommendations - Including treatment recommendations, lifestyle changes and links to doctors consultation based on predictions of the disease.
- Multimodal Health Monitoring Multimodal health monitoring is the combination of wearable devices and IoT sensors to interpret real-time health parameters and symptoms to make more precise predictions.
- Increased Security and Privacy- The use of secure data storage and healthcare data standard compliance to guarantee user confidentiality.

REFERENCES

[1] Sogandi, F. (2024). Identifying diseases symptoms and general rules using supervised and unsupervised machine learning. *Scientific Reports, 14*, Article 17956. https://doi.org/10.1038/s41598-024-69029-8 Nature

[2] Dinsa, E. F., Das, M., & Abebe, T. U. (2024). AI-based disease category prediction model using symptoms from low-resource Ethiopian language: Afaan Oromo text. *Scientific Reports, 14*(1), 11233. https://doi.org/10.1038/s41598-024-62278-7 PubMed

[3] Ahsan, M. M., & Siddique, Z. (2021). Machine-learning-based disease diagnosis: A comprehensive review. *arXiv*. https://arxiv.org/abs/2112.15538 arXiv

[4] Tirkey, A., & Borah, R. J. (2025). Symptom-based disease prediction using machine learning: A web application approach. *International Journal of Pure and Applied Mathematics and Humanities*. [Manuscript in press]. https://www.researchgate.net/publication/385162172_Symptom-Based_Disease_Prediction_Using_Machine_Learning_A_Web_Application_Approach ResearchGate

[5] Sogandi, F., et al. (2024). Symptom-based machine learning models for disease prediction using Apriori and classification algorithms. *Scientific Reports, 14*, Article 17956. https://doi.org/10.1038/s41598-024-69029-8 Nature *(Note: This is the same as reference 1, but emphasizing algorithmic detail.)*

[6] S. M, P. N G, S. A. N and J. B, "The Disease Prediction System based on Symptom Identification and Image Processing," 2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE), Gautam Buddha Nagar, India, 2024, pp. 1075-1079, doi: 10.1109/IC3SE62002.2024.10593340.

[7] Eur. J. Med. Res. (2025). Unveiling the potential of artificial intelligence in revolutionizing disease prediction and diagnosis: A comprehensive review. *European Journal of Medical Research*. https://doi.org/10.1186/s40001-025-02680-7 BioMed Central

[8] Md Manjurul Ahsan & Zahed Siddique. (2021). Machine learning-based heart disease diagnosis: A systematic literature review. *arXiv*. https://arxiv.org/abs/2112.06459 arXiv