INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT (IJSREM)
VOLUME: 09 ISSUE: 01 | JAN - 2025 SJIF RATING: 8.448 ISSN: 2582-3930

Patient Sickness Prediction Using Machine Learning

Dhnyanesh Chaudhari[1] Vishakha Gedam[2]
dhnyaneshpc@gmail.com vishakhagedam@dypic.in

[1]UG Student, Department of Artificial Intelligence and Data Science, Engineering Ajeenkya DY Patil School of Engineering Lohegaon, Pune, Maharashtra India-412 105.
[2]Assistant Professor, Department of Artificial Intelligence and Data Science Engineering Ajeenkya DY Patil School of Engineering Lohegaon, Pune, Maharashtra India-412 105.

Abstract

The ability to predict a patient's sickness is essential to modern healthcare since it allows for timely treatments that can significantly improve patient outcomes. This work aims to predict the risk of specific illnesses using machine learning (ML) techniques based on large patient data. This project intends to build prediction models utilizing test results, lifestyle factors, and electronic health records (EHRs) to assist healthcare professionals in identifying at-risk patients and facilitating preventative healthcare treatments. The results of the study will enhance clinical judgment and patient care by applying efficient predictive analytics in the medical field.

Keywords: At-Risk Patients, Disease Prediction, Electronic Health Records (EHRs), Machine Learning (ML)

I. INTRODUCTION

Machine learning (ML) is the process of using historical data or example data to train computers to perform better. It includes the study of computer systems that learn from data and prior experiences to operate better. Training and testing are the two primary stages of the machine learning process. Machine learning has been crucial in the last several decades for making illness predictions based on a patient's medical history and symptoms. The use of machine learning in healthcare offers a strong foundation for effectively managing a range of medical conditions. ML technology help create prediction models that allow for quick data analysis and prompt outcomes by evaluating extensive healthcare data. In the end, this skill improves the quality of patient care by enabling medical practitioners to make well-informed judgments about patient diagnoses and treatment alternatives. One of the best examples of how machine learning is affecting the medical industry is the healthcare sector. This effort will concentrate on evaluating written and unstructured data in order to increase prediction accuracy utilizing massive datasets. To forecast diseases, the study will use algorithms like Random Forests, Naive Bayes, and Decision Trees. Today's healthcare systems are overloaded with enormous amounts of patient data produced by a variety of sources, such as laboratory testing, patient-reported outcomes, medical imaging, and electronic health records (EHRs). Detailed medical histories, demographic information (age, gender, ethnicity), genetic data, and lifestyle variables (diet, physical activity, smoking history) are just a few of the many types of information included in this data.Predicting patient illness accurately is essential because it allows medical professionals to take early action, which may stop illnesses from getting worse. Early identification enables prompt therapy modifications and individualized care regimens, which can greatly improve patient outcomes. Furthermore, successful prediction models can result in more effective resource allocation within healthcare systems, guaranteeing that facilities and medical personnel are better equipped to meet patient demands. This can minimize needless hospitalizations and improve the quality of treatment.

Machine learning offers a powerful framework for developing sophisticated predictive models that analyze complex datasets and uncover hidden patterns within the data. By leveraging techniques such as supervised learning, where models are trained on labeled data, and unsupervised learning, which identifies patterns without predefined labels, machine learning can enhance clinical decision-making. These models can assist healthcare providers in identifying at-risk patients, forecasting disease outbreaks, and tailoring interventions based on individual patient profiles, ultimately transforming the landscape of healthcare delivery.

II. OBJECTIVE

The project's goal is to use a variety of patient data from electronic health records to create a machine learning model that can forecast patient ailments. It will include gathering and analyzing this data, evaluating various machine learning algorithms to identify the most effective ones for illness prediction, and developing ideal models in order to ensure accuracy. The study will assess the models' effectiveness using metrics like accuracy and recall in addition to looking into how the models may help medical professionals make informed decisions. We'll also discuss ethical concerns including following rules and safeguarding patient privacy. Improving sickness prediction is the ultimate goal in order to enhance healthcare results.

III. LITERATURE SURVEY

Several machine learning methods and algorithms that are useful in healthcare facilities have been used in several research to investigate illness prediction. This essay summarizes the methods and findings of a number of important studies from the research literature. A system with 94.8% accuracy was created by Min Chen et al. [2] using CNN-UDRP, CNN-MDRP, Naive Bayes, K-Nearest Neighbor (KNN), and Decision Trees. A disease risk prediction system with an accuracy of 82% was proposed by Sayali Ambekar et al. [3] utilizing convolutional neural networks, Naive Bayes, and KNN.A fuzzy-based method for predicting diabetes and liver problems was presented by Naganna Chetty et al. [4], who achieved accuracies of 97.02% and 96.13%, respectively. Dahiwade Dhiraj. [5] developed a model with a 91% accuracy rate for predicting heart illness using KNN and CNN. Last but not least, Ankita Dewan et al. [8] created a hybrid classification system that used Naive Bayes, Decision Trees, and Neural Networks to predict heart disease with an accuracy of 87%.

IV. PROPOSED SYSTEM

This system uses a decision tree classifier to assess the model and forecast illnesses based on patient symptoms. End users who are worried about their health are the target audience. To make predictions, the system makes use of machine learning technologies, most especially the decision tree classifier algorithm. This system is called "AI Diagnoser." With features designed to both inform and enhance their well-being, it serves those who constantly worry about their health. The "Disease Predictor," which recognizes possible diseases based on reported symptoms, is one important aspect. Integration, forecasting, route analysis, and prediction analysis are some of the parameters that are included in the prediction process. The three primary components of the system are the Admin, Doctor, and Patient modules.

VOLUME: 09 ISSUE: 01 | JAN - 2025 SJIF RATING: 8.448 ISSN: 2582-3930

V. CONCEPTUAL ARCHITECTURE

A user's risk of having an illness is determined using machine learning-based disease prediction, which takes into account a variety of symptoms they report, including headaches, back discomfort, runny nose, and more. To match the user's symptoms and make illness diagnosis easier, the system makes use of a variety of databases. Prior to being arranged using classification techniques, these datasets are first divided into more manageable, smaller chunks.

Machine learning systems then process this data by examining the previously specified user inputs. Following the patient's submission of information, the system aggregates the data and applies the illness prediction model to produce a prognosis about possible health problems.

.

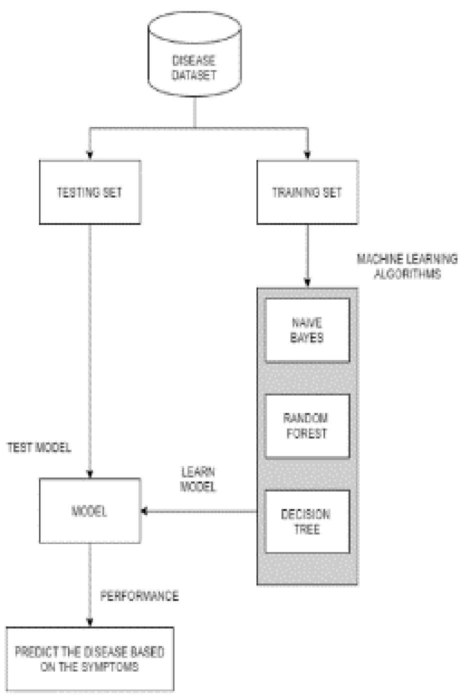


Fig Conceptual Architecture

VI. BLOCK DIAGRAM

User Roles: Three different user categories are supported by the system:

1. Patients: They may schedule online consultations with medical professionals, enter their symptoms, and get predictive insights. Because of its straightforward design, the user interface is

simple to use.

2. Doctors: Medical personnel have access to patient symptom information, are able to diagnose patients, and may consult with patients. On the basis of the system's predictions, they may also provide customized feedback.

3. Administrators: To guarantee a seamless user experience, administrators control user access, preserve data integrity, and keep an eye on system performance.

User Verification Procedure:

- To verify each user's identity, the system has a comprehensive user authentication procedure. This security measure guarantees the confidentiality of consultations and protects patient information.

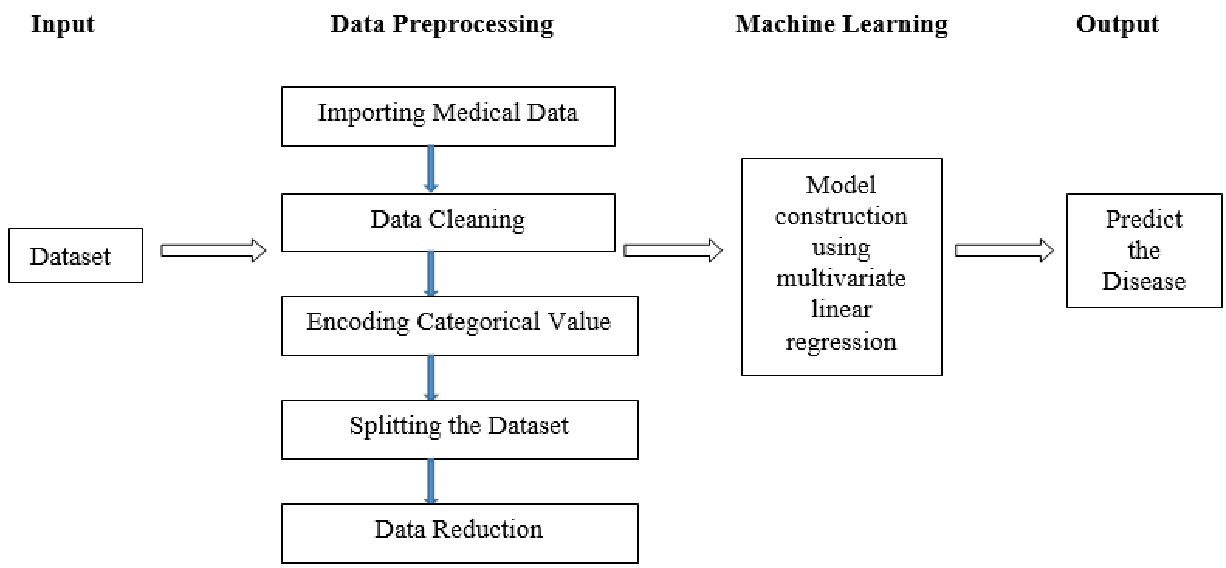


Fig :- Block Diagram

Symptom Entry and Illness Prediction:

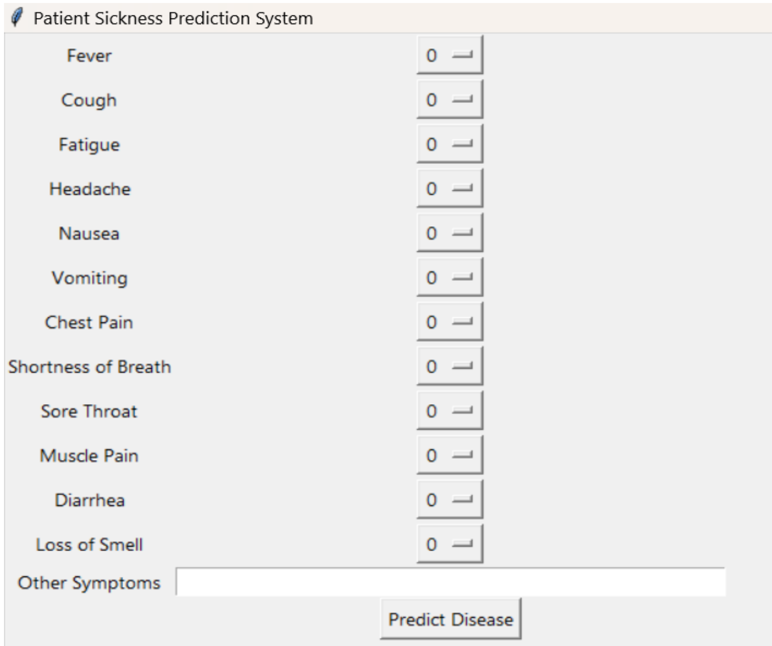
- The system allows users to enter their symptoms, analyzes the information, and makes predictions about possible health problems using machine learning algorithms. This predictive function promotes prompt medical consultation and helps individuals better comprehend their symptoms.

Role-Based Access Management: To guarantee that users' permissions are appropriate for their designated responsibilities, a role-based access control system is put in place. This improves critical data protection and facilitates efficient user access management.

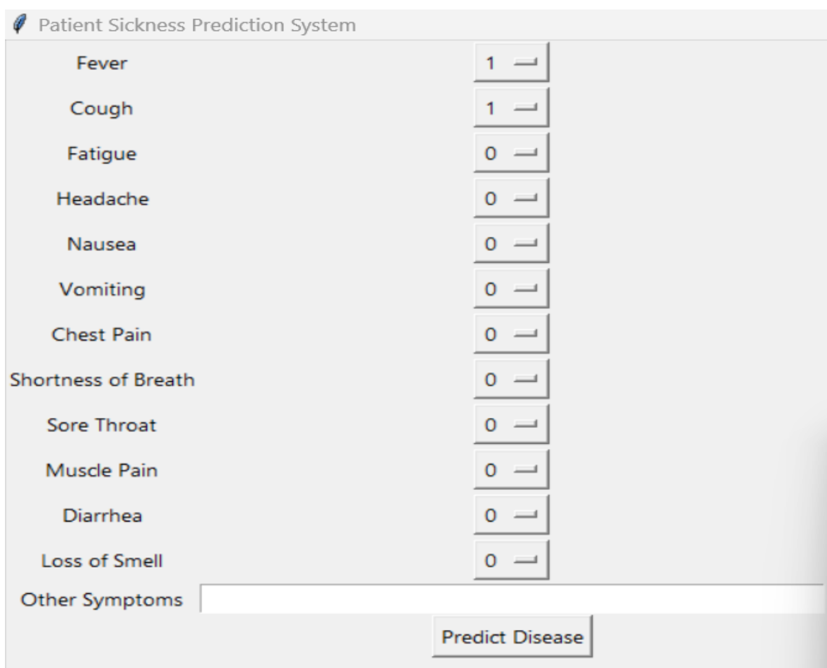
Online Consultation Capability: This feature allows patients to engage in real-time conversations with doctors by participating in online consultations. Those who live in remote places or may have trouble traveling would especially benefit from this.

Remote Consultations: Using the method, patients may communicate with medical professionals from the comfort of their own homes. This feature makes things more convenient by allowing people to get medical advice without having to travel.

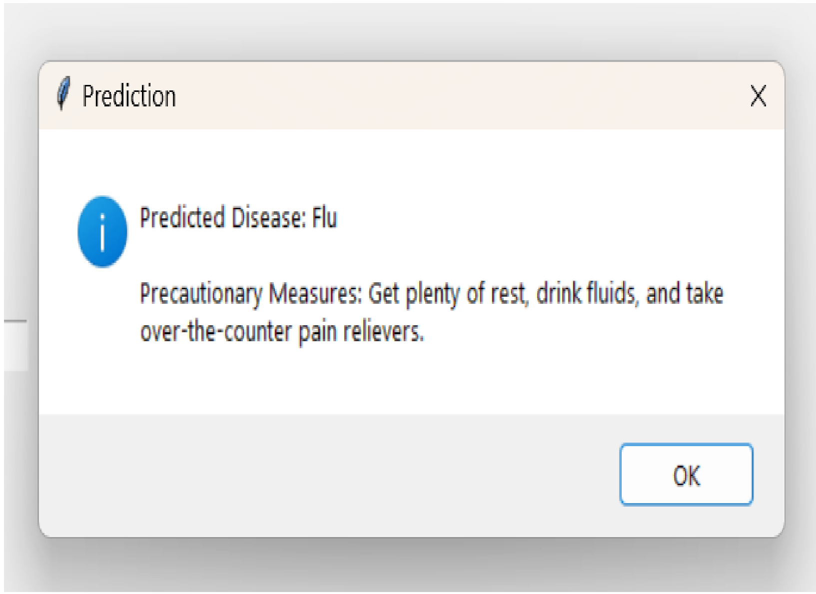
VII. SYSTEM PROTOTYPE



1. Home Screen



2. SELECTING SYMPTOMS



3. OUTPUT

VIII. ALGORITHMS USED

1) **Decision Trees**

• **Description:** Decision trees use a framework like a tree to make decisions, with nodes standing for features, branches for decision rules, and leaves for results.

• **Functionality**: To produce as homogeneous subsets as feasible, the algorithm divides the data according to particular criteria. Until a stopping requirement is satisfied, such reaching a maximum depth or obtaining a minimum number of samples in a node, this process keeps going.

• **Uses:** Decision trees are often used in medical contexts to diagnose illnesses based on reported symptoms since they are excellent at visualizing decision routes.

2) **Naive Bayes**

• **Description**: Based on the idea that characteristics are independent of one another, the Naive Bayes classification technique is based on probability theory.

• **Functionality**: Based on the symptoms, it determines the probability of each condition and chooses the one with the highest chance. Even with big datasets, it may function well because of the independence condition, which makes the computations simpler.

• **Applications**: Naive Bayes is frequently used for text categorization, but it may also be used to forecast illnesses by analyzing combinations of symptoms.

3) **K-Nearest Neighbors (KNN)**

• **Description**: KNN is a simple instance-based learning technique that uses the classes of the dataset's closest neighbors to categorize each instance.

• **Functionality**: KNN gives the most prevalent diagnosis among the k most comparable previous instances to the new instance when a patient presents their symptoms.

• **Uses:** This technique is useful in situations where historical instances might offer information about possible illnesses by comparing symptoms.

4) **Support Vector Machines (SVM)**

• **Description**: SVM is a supervised learning method that divides classes by creating a hyperplane in a high-dimensional space.

• **Functionality**: SVM works especially well in complicated datasets by optimizing the distance between classes, which reduces the likelihood of overfitting.

• **Uses**: SVM works well for problems involving the prediction of diseases in which symptoms are represented in high-dimensional feature spaces, including complicated health data or genetic data.

5) **Random Forest**

• **Description:** To increase accuracy, Random Forest is an ensemble technique that constructs many decision trees and combines their results.

• **Functionality**: It improves overall predictive performance and reduces the danger of overfitting by averaging predictions from many trees.

• **Applications**: Because of its resilience and capacity to efficiently handle a large number of features, this algorithm is frequently employed in illness prediction.

6) **Logistic Regression**

• **Description:** The statistical method known as logistic regression is primarily applied to tasks involving binary categorization.

• **Functionality**: It uses a logistic function to estimate probabilities in order to represent the link between a dependent variable (such as the existence of a disease) and independent variables (such as symptoms).

• **Uses**: Based on a collection of symptoms, this approach is very helpful in determining whether a disease is present.

IX. CONCLUSION

Healthcare technology has advanced significantly with the use of machine learning (ML) in patient illness prediction. Healthcare professionals may more accurately forecast possible diseases by utilizing advanced algorithms to evaluate enormous volumes of patient data, such as symptoms, medical history, and lifestyle variables. In summary, the use of machine learning to forecast patient illness has enormous potential to transform the healthcare system and make it more effective, accessible, and sensitive to the requirements of each patient. For these predictive models to be improved and smoothly incorporated into clinical practice, more research and development in this field will be essential.

X. FUTURE SCOPE

Integration with Wearable Devices:

- The increasing use of wearable health technologies offers a chance to update prediction models using real-time vital sign and physical activity data. This might make it possible to monitor health continuously and make it easier to identify any medical problems early.

Advancements in Natural Language Processing (NLP):

- Progress in NLP technologies will allow for the analysis of unstructured data from electronic health records and patient narratives. This capability will enhance the depth of information utilized in predictions, improving diagnostic accuracy.

Expansion of Telemedicine:

- The analysis of unstructured data from patient narratives and electronic health records will be possible thanks to advancements in NLP technology. This talent will improve• As telehealth services expand, including machine learning (ML)-driven predictive analytics into virtual consultations might enable medical professionals to make well-informed judgments based on past data and current patient-reported symptoms.

Development of Collaborative Healthcare Platforms:

- By enabling shared access to patient data and insights from predictive analytics, future technologies may allow healthcare practitioners to collaborate. Care coordination would improve as a result, especially for patients with chronic, complicated illnesses.

Establishment of Ethical and Regulatory Standards:

- In order to protect data privacy and guarantee responsible AI use, the growth of ML applications in healthcare will require the creation of ethical and legal frameworks. Establishing patient trust and encouraging broad adoption will need these concepts.

Empowerment and Engagement of Patients:

- By using predictive analytics to provide insights, future developments may concentrate on boosting patient involvement. Patients can participate more actively in their healthcare decisions if they are informed about the hazards to their health.

Validation in Real-World Settings:

- ML models must be continuously validated in clinical settings in order to increase their dependability and win over medical professionals. The goal of future studies is to improve these models so they work well for a range of demographics.

Interdisciplinary Approaches:

- Adding viewpoints from disciplines like public health, psychology, and sociology to machine learning models can promote a thorough comprehension of health concerns and result in solutions that tackle the intricate nature of patient care.

REFERENCE

[1] M. Jiang, Y. Chen, M. Liu, S. T. Rosenbloom, S. Mani, J. C. Denny, and H. Xu, "A study of machine-learning-based approaches to extract clinical entities and their assertions from discharge summaries," J. Am Med Inform Assoc, vol. 18, no. 5, pp. 601–606, 2011.

[2] Deeraj Shetty, Kishor Rit, Sohail Shaikh and Nikita Patil, " Diabetes DiseasePrediction Using Data Mining" IEEE, 978-1-5090-3294-5/17, 2017.

[3] Lambodar Jena and Ramakrushna Swain, "ChronicDisease Risk Prediction using Distributed Machine Learning Classifiers" IEEE, 978-1-5386-2924-6/17, pp. 170-173, 2017.

[4] Dhiraj Dahiwade, Gajanan Patle and Ektaa Meshram, "Designing Disease Prediction Model Using Machine Learning Approach" IEEE Xplore Part Number: CFP19K25-ART; ISBN: 978-1-5386-7808-4, pp.1211-1215, 2019.

[5] Naganna Chetty, Kunwar Singh Vaisla and Nagamma Patil, "An Improved Method for Disease Predictionusing Fuzzy Approach" IEEE,DOI 10.1109/ICACCE.2015.67, pp. 569-572, 2015.

[6] Dhomse Kanchan B. and Mahale Kishor M., "Study of Machine Learning Algorithms for Special Disease Prediction using Principal ofComponent Analysis" IEEE, 978-1-5090-0467-6/16, pp. 5-10, 2016.

[7] Ankita Dewan and Meghna Sharma, "Prediction of Heart Disease Using a Hybrid Technique in Data Mining Classification" IEEE, 978-9-3805-4416-8/15, pp. 704-706, 2015.

[8]Pahulpreet Singh Kohli and Shriya Arora, "Application of MachineLearning in Disease Prediction" IEEE, 978-1-5386-6947-1/18, pp. 1-4,2018.

[9] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang,"Disease prediction by machine learning over big data from healthcare communities" IEEE Access, vol. 5, no.1, pp.8869–8879, 2017.

10] Sayali Ambekar, Rashmi Phalnikar, "Disease RiskPrediction by Using Convolutional Neural Network" IEEE, 978-1-5386-5257-2/18,2018.