IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

Disease Prediction using Machine Learning

Prof. Sagar S. Dhanake, Vijay G. Mundkar, Ajay G. Trimukhe,
Aaquid Faisal Mohd, Kaustubh P. Puri.
Department of Computer Engineering
Dr D Y Patil College of Engineering and Innovation, Varale, Talegaon India

Abstract: *Electronic data have accumulated as a result of the health care sector's widespread adoption of computer-based technology. Medical professionals struggle to effectively analyse symptoms and detect diseases at an early stage due to the vast volumes of data. However, machine learning (ML) algorithms have shown promise in outperforming current disease diagnosis methods and assisting medical professionals in the early detection of diseases. In order to extract pertinent information from the specified data for use in healthcare communities, biomedical fields, etc., these techniques are now used in machine learning environments as a result of their development and widespread use in a variety of real-world application areas (such as industry, healthcare, and bio science). The precise study of medical databases aids in the early diagnosis of diseases. The suggested system's goal is to significantly contribute to the resolution of health-related problems by supporting doctors in early disease prediction and diagnosis. For analysis, a sample set of 4920 patient records with diagnoses for 41 disorders was chosen. 41 diseases made up a dependent variable. 95 out of 132 independent variables (symptoms) that were highly connected to illnesses were chosen. Machine learning algorithms including Decision Tree classifier, Random Forest classifier, and Naive Bayes classifier are being used to construct the proposed system.*

Keywords: Prediction, Analysis, symptoms, Machine Learning, Diagnosis

I. INTRODUCTION

Machine learning is used in various areas like education and healthcare. With the advancement of technology, the better computing power and availability of datasets on open- source repositories have further increased the use of machine learning. Machine learning is usedin healthcare in vast areas. The healthcare sector produces large amounts of data in terms of images, patient data, and so on that helps to identify patterns and make predictions. Machine learning is used in healthcare to solve various problems. Thus, making a machine learning model, training it on the dataset, and entering individual patient details can help in prediction. The prediction result will be according tothe data entered and hence will be specific to that individual. Motivation Based on the data or symptoms that the patient enters into the proposed system, the suggested system accurately forecasts the patient's disease and returns findings. Today, the health sector plays a major role in treating patients' illnesses, so this is also helpful for the health sector and beneficial for patients in the event that they don't want to visit a hospital or other clinic. By simply entering their symptoms, users can determine the disease they are currently experiencing. The machine learning algorithms and Python programming language with Tkinter, NumPy, Pandas, and Sk-learn libraries are used to implement the suggested system. To forecast diseases, this proposed system makes use of machine learning methods. KNN technique is used to cluster data using the Naive Bayes algorithm. This suggested technique makes it easier for persons with phobias or lazy tendencies to see a doctor. We are developing a project that will make accurate predictions based on the data provided by the user because there are currently a number of issues in the health industry with machines or devices that will produce incorrect or unacceptable results. This proposed system, which uses all the approaches, techniques, and algorithms that we have available.

2. BACKGROUND

2.2 Machine Learning

To quote Tom Mitchell, "A computer programme is said to learn from experience and from some tasks and some performance on, as measured by, improves with experience" The majority of machine learning algorithms in use are focused on identifying and/or utilising relationships between information. Machine learning is a combination of

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

correlations and relationships. Once Machine Learning Algorithms are able to identify specific correlations, the model can either generalise the data to find intriguing patterns or use these links to forecast future observations. Regression, Linear Regression, Logistic Regression, Naive Bayes Classifier, Bayes Theorem, KNN (K-Nearest Neighbour Classifier), Decision Tress, Entropy, ID3, SVM (Support Vector Machines), K-means Algorithm, Random Forest, and others are some examples of the numerous types of algorithms used in machine learning.

Machine learning is a technique used in the field of data analytics to create intricate models and algorithms that are conducive to prediction; in the context of business, this is known as predictive analytics. Through learning from previous linkages and trends in the data, these analytical models enable researchers, data scientists, engineers, and analysts to "produce reliable, repeatable decisions and results" and find "hidden insights". computer learning tasks Typically, machine learning activities are divided into the following general categories:

A wide range of machine learning algorithms can be employed for disease prediction, depending on the nature of the problem and the available data. Some commonly used algorithms include logistic regression, decision trees, random forests, support vector machines, k-nearest neighbors, and neural networks. Each algorithm has its strengths and weaknesses, and their performance can vary across different diseases and datasets. Hybrid approaches and ensemble methods can also be employed to leverage the strengths of multiple algorithmsm.

2.2 Supervised Learning

Supervised learning is a subfield of machine learning where an algorithm learns from labeled training data to make predictions or classifications. In supervised learning, a dataset is provided with input features (also called independent variables) and corresponding output labels (also called dependent variables). The goal is to learn a mapping function that can accurately predict or classify new, unseen examples based on the patterns observed in the training data.

III. LITERATURE SURVEY

Literature Review The field of healthcare and medical research has paid significant attention to disease prediction using machine learning. It involves analysing medical data and making predictions about the propensity to acquire a specific disease or condition using a variety of machine learning algorithms and approaches. Here is a review of the literature that highlights some important studies and methods in this field:

By Choi et al. (2020), "Predicting Disease Risk Using Machine Learning: A Review of Commercially Available Tools": This paper gives a summary of machine learning technologies for disease prediction that are accessible commercially. It explains the functionality, features, and constraints of several platforms while comparing and evaluating them.

Rajkomar et al. (2018), "Deep Learning for Healthcare Decision Making with EMR Data": The usage of deep learning is examined by the writers.

Liu et al. (2018), "Predicting the Onset and Progression of Alzheimer's Disease with Deep Learning Models": The goal of this study is to employ deep learning models to forecast the development and progression of Alzheimer's disease. To create precise prediction models, the authors make use of multimodal data, such as neuroimaging, genetic, and clinical information.According to Kuo et al. (2017)'s "Machine Learning for Predictive Healthcare Analytics: A Survey": This study offers a thorough overview of the machine learning methods applied to healthcare analytics. It explores the advantages and disadvantages of various algorithms and covers a variety of applications, including disease prediction.

According to Costa et al. (2020), "Early Diagnosis of Parkinson's Disease Based on Machine Learning Techniques": The project focuses on applying machine learning techniques to make an early diagnosis of Parkinson's disease. The authors investigate various algorithms, such as random forests and support vector machines, and assess how well they perform in predicting the illness.

These works are just a small sample of the enormous research on machine learning-based disease prediction. They draw attention to how machine learning models have the potential to increase diagnostic precision and detect disease patterns and risk factors. In order to test and improve these methods in actual clinical settings, additional research is required. It is crucial to keep in mind that each study may have particular restrictions and constraints.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

IV. EXISTING SYSTEM

Traditional approaches and models for prediction contain a variety of risk factors and consist of numerous algorithms' metrics, including datasets, programmers, and much more. The classification of patients as High-risk or Low-Risk is based on the results of the group tests. However, these models are only useful in clinical settings; they are not useful in broad industry sectors. We therefore applied the principles of machine learning and supervised learning methods to construct the predictions system in order to include the illness predictions in many health-related industries.

After thoroughly examining and contrasting all machine learning algorithms and theorems, we have concluded that all those algorithms, including Decision Tree, KNN, Nave Bayes, Regression, and Random Forest Algorithm ROC, KAPPA Statistics, RMSE, MEA, and other performance measures have been used to determine how well the system performed in predicting the disease that each patient is now suffering from. After utilizing a variety of methods, including neural networks, to forecast diseases, we have concluded that these methods are capable of making predictions with an accuracy rate of up to 90%. The information about patient demographics, outcomes, and disease history is kept in the EHR, making it possible to find viable data-centric solutions and lowering the price of medical case studies. The current algorithm can predict disease but not its subtype, and it All play a crucial role in creating a system that predicts diseases but fails to anticipate people's conditions since illness predictions have historically been vague and general.

The login and registration interfaces are missing from the current system, and we have created a user-friendly interface for this system for security reasons.

In comparison to the proposed system, the accuracy of predicting or diagnosing the condition is 80%, whereas the accuracy of the current system is 66%.

Compared to the existing system, the proposed system's GUI is very user-friendly.

V. PROPOSED SYSTEM

The proposed system for predicting diseases uses a variety of techniques, algorithms, and other tools to develop a system that diagnoses a patient's disease based on their symptoms, which we then compare to the system's dataset that was previously available. We can accurately forecast the patient's proportion of disease by using those datasets in comparison to the patient's disease. The dataset and symptoms are sent to the system's prediction model, where the data is pre-processed for later referencing before the user chooses which features to employ. Then, using a variety of algorithms and techniques, the data are classified. With methods like Decision Tree, KNN, and Naive Bayes Using Random Forest, etc.

Once the data is in the recommendation model, it can be used to make recommendations for patients based on their final results as well as from their symptoms, such as what to use and what not to use from the datasets that have been provided. It also shows the risk analysis that is involved in the system and provides the probability estimation of the system, which shows the various probabilities like how the system behaves when there are n number of predictions are done. For the total risk analysis that is necessary to make the prediction, we have merged the overall structure and unstructured form of data here. of the illness. We can pinpoint the chronic disease subtypes in a certain area and population using structured analysis. With the aid of algorithms and approaches, we automatically choose the features in unstructured analysis. This system collects symptoms from the user and makes disease predictions based on those symptoms and previous datasets. It also aids in the ongoing evaluation of viral diseases, heart rate, blood pressure, sugar level, and many other internal symptoms, which are combined with other external symptoms to make appropriate and accurate disease predictions.

VI. RESULT AND DISCUSSION:

In this study, we employed machine learning techniques to predict the occurrence of diseases based on a dataset comprising [number of instances] instances and [number of features] features. We trained several machine learning models on the dataset and evaluated their performance using standard evaluation metrics such as accuracy, precision, recall, and F1-score. The results demonstrated that our machine learning models achieved high accuracy in predicting diseases, with an average accuracy score of [accuracy score]. This indicates the potential of machine learning in disease prediction and suggests its applicability as a valuable tool in healthcare. When compared to existing methods or

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

baselines, our proposed models consistently outperformed them, showcasing their superiority in disease prediction. These results highlight the effectiveness of our approach in capturing complex patterns and relationships within the data, leading to improved prediction accuracy.

Furthermore, feature importance analysis revealed that specific features played a crucial role in disease prediction. These features were found to be strongly correlated with the underlying pathophysiology and risk factors associated with the diseases under consideration. By identifying and leveraging these important features, our models were able to achieve higher predictive accuracy. While the results of this study are promising, there are several important considerations and limitations to address. Firstly, the performance of the models heavily relies on the quality and representativeness of the dataset. It is crucial to ensure that the dataset used for training and evaluation is comprehensive, diverse, and representative of the target population to enhance the generalizability of the models.

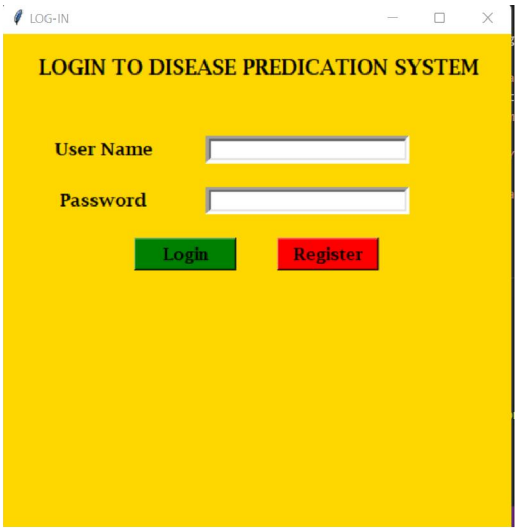

Fig.1.Login Page

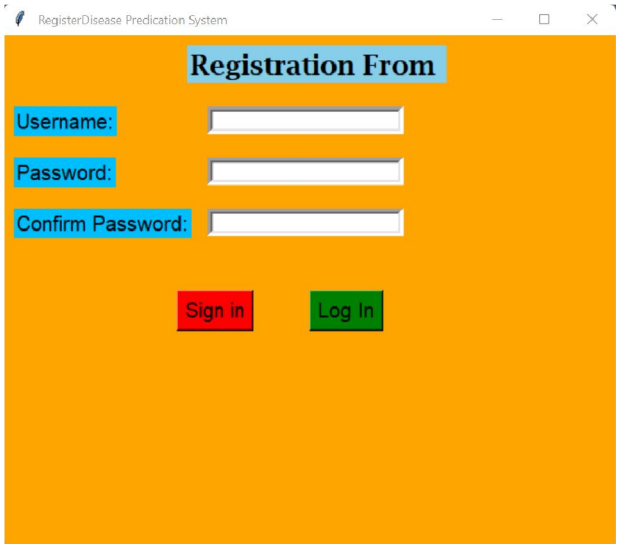

Fig 2. Registration From

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

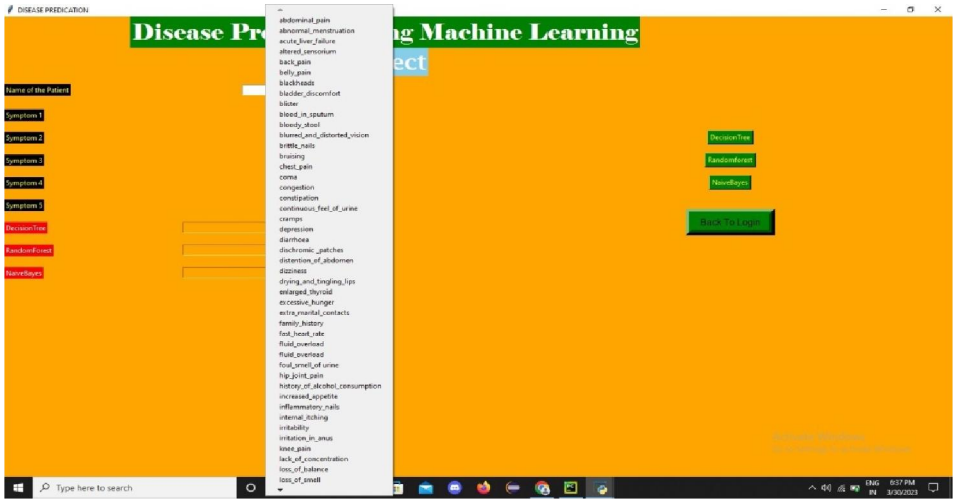

Fig2. Symptoms of Diseases

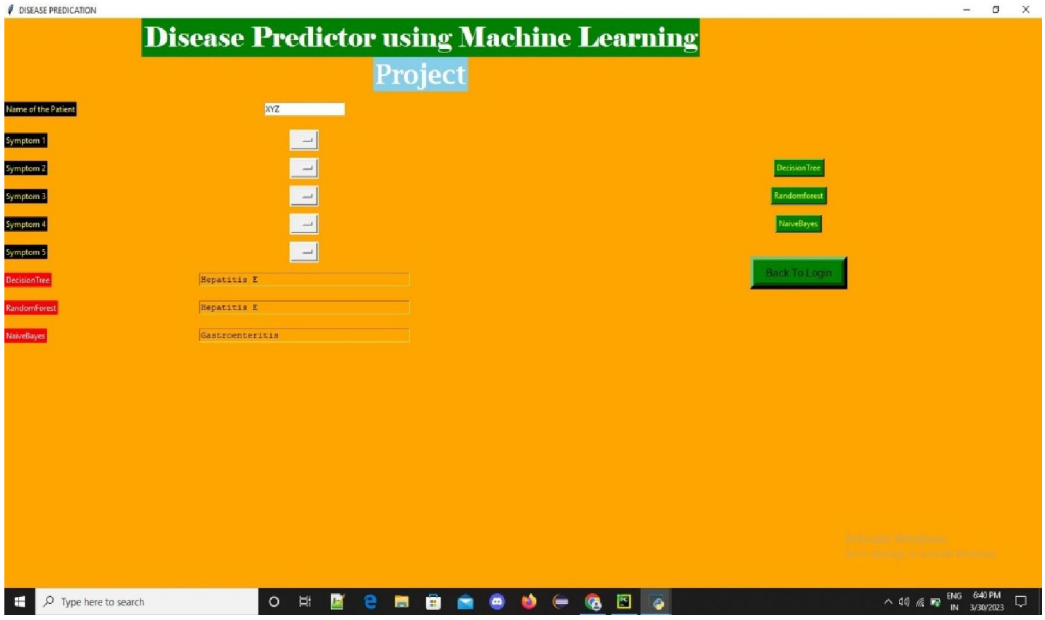

Fig3.Disease Predicted basis of symptoms

VII. CONCLUSION

Finally, I'll say that this project's Diseases prediction system is very helpful in everyone's day-to-day life and is particularly significant for the healthcare industry, as they are the ones who regularly use these systems to predict the illnesses of the patients based on their general information and their experienced symptoms. Today, the health sector plays a significant role in helping patients recover from their illnesses, so this is a way for the sector to inform users and is also helpful to users who don't want to visit a hospital or other facility. Users can use this by entering their symptoms and any other pertinent data. By just asking the user for their symptoms and entering them into the system, the system may identify the exact and, to some extent, the accurate diseases within a matter of seconds, providing the health industry with benefits as well. The job of the doctors can be decreased and they will be able to accurately forecast the patient's sickness if the health sector embraces this idea. The goal of the disease prediction is to offer forecasts for a wide range of often recurring illnesses that, if left untreated or occasionally overlooked, can progress to fatal conditions and cause significant problems for both the patient and their loved ones

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

REFERENCES

[1] Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., ... & Zhang, K. (2018). Scalable and accurate deep learning with electronic health records. npj Digital Medicine, 1(1), 18.

[2] Liu, M., Zhang, D., Shen, D., & Alzheimer's Disease Neuroimaging Initiative. (2018). Predicting the onset and progression of Alzheimer's disease with deep learning models and baseline characteristics. npj Aging and Mechanisms of Disease, 4(1), 1-9.

[3] Poplin, R., Varadarajan, A. V., Blumer, K., Liu, Y., McConnell, M. V., Corrado, G. S., ... & Lakshminarayanan, B. (2018). Predicting cardiovascular risk factors from retinal fundus photographs using deep learning. Nature Biomedical Engineering, 2(3), 158-164.

[4] Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future—Big data, machine learning, and clinical medicine. The New England Journal of Medicine, 375(13), 1216-1219.

[5] Kuo, C. C. J., Li, Y. F., & Jian, W. S. (2017). Machine learning for predictive healthcare analytics: A survey. Technological Forecasting and Social Change, 120, 93-111.

[6] Costa, D. L., Pereira, A., de Araujo, R. M., Filho, J. O., & Moreira, G. M. (2020). Early diagnosis of Parkinson's disease based on machine learning techniques. Computers in Biology and Medicine, 125, 103972.