International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 04 | Apr 2024 www.irjet.net p-ISSN: 2395-0072

MACHINE LEARNING APPROACHES TO MULTI-DISEASE PROGANASTICATION

[1]G K Karthik, [2]S Sahishnu Nag, [3]K Jwalitha

[1,2,3] Student, GITAM (Deemed to be University), Visakhapatnam, Andhra Pradesh, India.

---

1. Abstract -

Progress in machine learning in the biomedical and healthcare domains has made accurate analysis of medical data better for early sickness detection, patient treatment, and community services. Insufficient quality of medical data results in a decline in the study's accuracy. Furthermore, distinct geographical areas display distinct manifestations of certain localised illnesses, thus undermining the forecasting of disease epidemics. The suggested approach offers machine learning algorithms for accurate forecasting of different illness incidences in communities where diseases are common. It tests the modified estimate models using actual hospital data that has been gathered. It uses a latent factor model to reconstruct the missing data in order to get over the challenge of incomplete data. It tests a localised chronic form of cerebral infarction. It makes use of both structured and unstructured hospital data. By mining data sets for conditions like diabetes, breast cancer, and heart disease, it makes predictions about likely ailments. To the best of our knowledge, no previous effort in the field of medical big data analytics has taken into account both forms of data. Our suggested technique outperforms numerous common estimate algorithms in terms of calculation accuracy, reaching 94.8%, and has a faster rate of convergence than the machine learning disease risk prediction algorithm.

Key Words: Heart Disease prediction, Diabetes, Breast Cancer.

2. Introduction

In today's digital age, data has become a valuable asset, with vast amounts being generated across various industries. The healthcare sector, in particular, relies on patient-related information, collectively referred to as healthcare data. In this context, we propose a general architecture for predicting illnesses in the healthcare sector. Many existing models tend to focus on analyzing one disease at a time, such as diabetes, cancer, or skin conditions. However, there is a noticeable absence of a comprehensive system that can simultaneously assess multiple diseases. Our primary objective is to provide customers with accurate and real-time disease predictions, along with corresponding information about the associated symptoms.

To address this challenge, we are introducing a Django-based system designed for predicting specific medical conditions. In our initial implementation, we will focus on analyzing malaria, heart disease, and diabetes. However, it's important to note that we have the flexibility to expand the range of diseases in the future. Our approach combines Django, a robust web framework, with machine learning techniques to develop multiple disease prediction models.

One notable advantage of our system is its comprehensive consideration of all relevant factors contributing to each disease during the analysis, which enables more precise and effective disease identification. To ensure the preservation of the model's behavior, we utilize Python pickling, allowing us to save and load the trained model as a pickle file in Python. This approach ensures that our system maintains its predictive accuracy and functionality.

Many existing healthcare systems have been primarily designed to assess individual diseases separately. For instance, one system might be dedicated to diabetes, another to diabetic retinopathy, and yet another to heart disease prediction. This fragmented approach often requires organizations to deploy a variety of models to evaluate patient health information. These systems are tailored for analyzing specific ailments in isolation.

In contrast, our proposed system offers a more versatile approach. Users of our multi-disease prediction system can conveniently assess several diseases on a single webpage. They no longer need to navigate multiple platforms to determine whether they might be affected by a particular illness. To utilize this comprehensive disease prediction system, users simply need to select the name of the disease of interest, input the relevant parameters, and click the "submit" button. The system will then invoke the appropriate machine learning model to forecast the outcome and display the results on the screen. This integrated approach streamlines the process for users, providing a more user-friendly and efficient experience.

The need for accuracy and intelligence is rising behind the purplish patch of technology that the Earth is currently experiencing. The folks of today probably have an internet addiction, but they don't give a damn about their physical well-being. Individuals neglect minor issues and avoid going to the hospital, which over time can develop into more serious illnesses. Our primary goal is to create a system that, by using the rapidly advancing technology, can diagnose various illnesses based on the symptoms stated by patients, saving them trips to the doctor or hospital.

3. Literature Review

3.1. The study emphasizes the significant impact of diabetes as one of the most perilous diseases worldwide.With the potential to lead to various health complications, including vision impairment. Given the ease and adaptability of machine learning techniques in predicting a patient's health status, these methods were employed in the research to detect the presence of diabetes accurately. The primary objective of this investigation was to create a diagnostic system that enables patients to precisely identify diabetes.

The research assessed the performance of four key algorithms: Decision Tree, Naïve Bayes, and Support Vector Machine (SVM), with accuracy rates of 85%, 77%, and 77.3%, respectively. In addition to the training phase, the researchers also implemented an Artificial Neural Network (ANN) algorithm to observe the network's responses, determining whether the disease was correctly identified or not. The study further compared each model's accuracy, precision, recall, and F1 score support.

3.2. Illustrating the heart's vital importance to all living things is the paper's principal goal.
The Identification and prediction of heart-related illnesses is critical due to their potential for fatality. In this way, artificial intelligence and machine learning aid in the prediction of many natural disasters. To calculate the accuracy of machine learning for heart disease prediction, the authors of this work use k-nearest neighbor, decision tree, linear regression, and SVM, training and testing on the UCI repositor dataset. They also evaluated the algorithm's accuracy against that of the K-Nearest Neighbor (87%), Support Vector Machine (83%), Decision Tree (79%), and Linear Regression (78%).

3.3. A personalised and economical approach to the identification of Alzheimer's disease using machine learning:-
In early stages of Alzheimer's disease, cognitive impairment may be minimal, making a diagnosis challenging. However, streamlining the diagnosing process would be extremely beneficial, as this is the window of time when treatment is most likely to be successful. A machine learning method for an affordable, individualized AD diagnosis is described and evaluated. It computes the sequence of biomarkers that is most useful or economical to identify patients by using locally weighted learning to create a classifier model for each patient. By comparing patients with AD to controls and patients with MCI who developed AD within a year to those who did not, we were able to classify AD patients using ADNI data. Although the method's performance was equivalent to that of analyzing all the data at once, it greatly decreased the quantity (and expense) of biomarkers required to verify a patient's diagnosis. As a result, it might be helpful in therapeutic settings and aid in the precise and tailored detection of AD.

3.4. Quantized Analysis Using Cardiac Sound Characteristic Waveform Method for Heart Valve Disease:-
A novel quantization diagnosis approach, known as the cardiac sound characteristic waveform, was presented to analyse four clinical heart valve sounds in order to precisely and efficiently analyse heart valve illness. To gather signal, a BIOPAC acquisition system was employed. Through Ethernet, the data from the user is sent for use on PC real-time display, evaluation, and storage. A one degree of independence analytical model is created to takeout the distinctive phase diagram. Additionally, parameters for diagnostics were computed and distinguished between normal Heart valve illness and cardiac sound using an understandable diagrammatic depiction, making it simple for even a novice user to track the progression of their pathology. Lastly, a case study showing the patient's condition before and after surgery is presented to confirm the effectiveness and utility of the suggested approach.

4. Problem Identification and Objectives

Nowadays, a lot of machine learning models used in health care analysis focus on one illness at a time. For instance, the first is for study of the liver, while the others are for analysis of lung issues and cancer. To forecast more than one ailment, one must consult various websites. Under any common paradigm, more than one disease cannot be predicted by a single analysis. There could be a significant effect on patient treatment if certain models have lower

accuracy than others. An organisation must install many models in order to assess patient health records, which adds to the expense and effort involved. Very few parameters are taken into account by some of the current systems, which can lead to inaccurate findings.

5. Aim of the Project

When the quality of the medical data is inadequate, the analysis's accuracy is decreased.

Furthermore, distinct locations display distinct features of specific localized illnesses, thus impairing the forecasting of disease outbreaks. Nonetheless, the majority of the previous research focused on structured data.

There are no suitable approaches for managing semi-structured and unstructured data.

Both structured and unstructured data will be taken into account by the suggested system.

The application of machine learning algorithms increases the accuracy of the analysis.

6. System Methodology

6.1 existing system

A machine can identify diseases, but it is unable to identify the subtypes of diseases that result from the incidence of a single disease. It is unable to foresee every scenario in which people might behave. Only structured data is handled by the current system. The prediction mechanism is unclear and has a wide scope. Numerous illness estimate classes have been developed and implemented in the recent past. The established organizations set up a combination of machine learning algorithms that are carefully accurate in predicting illnesses. Still, there are disparities in the restraint with the current systems. First of all, the current systems are more expensive, only accessible to the wealthy. Additionally, it increases even more when it comes to people. Secondly, the current guess systems lack he several subtypes of diseases or disorders brought on by a single bug. For example, if a group of people is predicted to have diabetes, it is likely that some of them would have a complicated risk for heart viruses as a result of their diabetes. The remaining plans are unable to predict every scenario in which the tolerant could be found.

6.2 Proposed System

The planned machine learning-based multiple disease prediction system compares a patient's symptoms to a previously available dataset in order to predict the patient's disease. This is accomplished through the use of algorithms the system, where the data is pre-processed for future references. Next, machine learning algorithms like logistic regression are used to classify such data. The data is then fed into the recommendation model, which displays the risk analysis that is part of the system and offers the system's probability estimation so that it can show different probabilities, such as how the system behaves when n number of predictions are made. It also generates recommendations for patients based on and other diverse tools. We can accurately forecast the patient's % disease by comparing those datasets with the patient's condition. The user chooses the features by entering or selecting the different symptoms. The dataset and symptoms are sent to the prediction model of their symptoms and final results, indicating which treatments to take and which ones to avoid based on the provided datasets and results. By analysing data sets related to diabetes, breast cancer, and heart disease, it makes predictions about probable illnesses. To the best of our knowledge, no previous effort in the field of medical big data analytics has taken into account both forms of data.

A. Creating models for machine learning:

In order to create the necessary and appropriate machine learning models, we must first comprehend the problem statement of our project. Next, we gathered data from a number of publicly available sources, including Kaggle and the UCI Machine Learning Repository. Both the quantity and quality of the data are crucial since they influence our model. Following that, we preprocessed the data to make sure the information was gathered in the correct format. As soon as possible, we examine the available data to eliminate duplicates and address any missing values. We also look for anomalies. To determine the link between the variables, we visualise the data. Here, we address the skewness and derive some insightful conclusions. Examine data that has been separated into training and testing sets. Of our data, 80% are used for training and 20% are used for testing. The most crucial stage in creating a reliable and accurate machine learning model is this one. To obtain the best results, we have experimented with a number of machine learning algorithms.

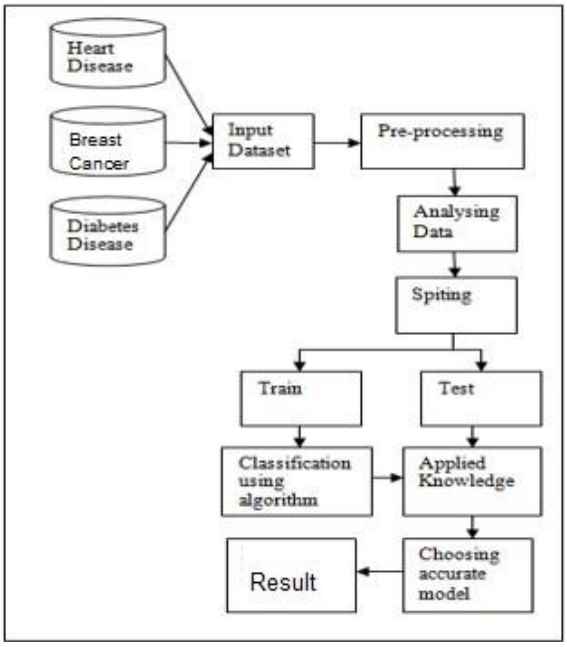

Fig1: Block diagram

B. Using Django to implement ML models:

The Django framework for Python is utilized to implement the machine learning models for various ailments. Django is a popular high-level language, open-source web framework for Python that is used to create websites and online applications. By offering an organized, reusable, and maintainable codebase, it was created with the intention of streamlining and expediting the web development process. Often referred to as Model-View-Template (MVT) in the Django context, the Model-View-Controller(MVC) architectural paradigm is followed by Django.

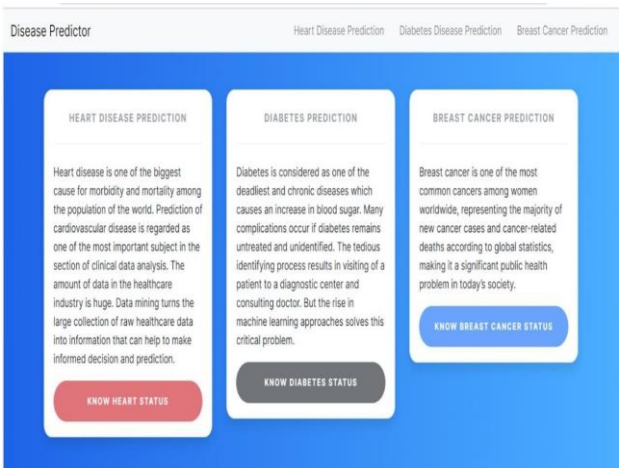

Fig 2: Graphical User Interface

7. Overview of technologies

Numpy

A well-liked open-source Python library for mathematical and numerical operations is called NumPy (Numerical Python). With the help of a number of mathematical functions, one can effectively work with sizable, multi-dimensional arrays and matrices. SciPy, pandas, scikit-learn, and many other libraries are built on top of NumPy, which is a core library in the Python data science and scientific computing ecosystem.

Pandas

Python programmers can use the Pandas library, which is an open-source data manipulation and analysis tool. Data science, machine learning, financial analysis, and other domains require it as a vital tool for data analysis, data transformation, and data cleaning since it offers user-friendly data structures and functions for handling structured data.

Scikit-learn

A machine learning library for the Python programming language, Scikit-learn, is sometimes shortened to sklearn. Many tools are available to carry out different machine learning tasks, such as dimensionality reduction, clustering, regression, and classification. NumPy and SciPy are just two of the well-known Python libraries upon which Scikit-learn is built. It is intended to be an intuitive, effective tool that is utilized extensively in both academic and commercial settings.

8. Testing:

System Testing

The testing of a Multi-Disease Prognostication system involved the application of various Machine Learning and Data Science methodologies. These approaches were instrumental in analyzing and predicting the progression of multiple diseases. The research focused on developing a thorough system capable of correct forecasting the progression of different illnesses simultaneously.

A variety of Data Science methods and Machine Learning algorithms were used in the testing stage. KNN and Random Forest were two of these. Through the application of these various approaches, the research team aimed to evaluate the prognostication system's efficacy and efficiency. To allow for a comprehensive assessment of the models' predictive skills, the study also assessed each model's accuracy, precision, recall, and F1 scores.

Categories of testing

8.1 Unit Testing

Unit testing, which entails testing individual programme units or components, is a crucial practise in software development. Unit testing is essential for guaranteeing the correctness and dependability of the produced models in the context of machine learning and data science approaches for multi-disease prognostication. In order to determine whether a function or method generates the desired result for a given input, it must be tested. Unit tests are usually created to evaluate a number of factors, such as model interpretability, prediction accuracy, data preprocessing, and training.

In order to evaluate the code's resilience, developers design test cases that cover a variety of situations and edge cases during unit testing. This procedure enhances the overall quality of the programmer by assisting in the early detection and correction of any possible problems during the development cycle. Maintaining the codebase and promoting teamwork among members also depend on comprehensive documentation of the tests and their outcomes.

8.2 Integration Testing

During the crucial integration testing stage of the software testing process, separate software modules or components are put together and tested collectively. Assuring that these interconnected components function flawlessly together with detecting and resolving any possible problems that can result from their interactions is the main objective of integration testing. In the software testing life cycle, integration testing usually comes before system testing and after unit testing.

8.3 System Testing

An essential step in the software testing process is system testing, which assesses the program as a whole. This kind of testing evaluates whether the software's modules and integrated components meet the required standards and function as a cohesive unit. The software development lifecycle includes system testing, which is usually carried out after unit and integration testing.

8.4 White-Box Testing

White box testing is a type of software testing where the tester is privy to the program's inner workings, structure, and language or at the very least, what it is meant to do. That is its goal. It is employed for testing regions that are inaccessible from a black box level.

9. Results

Project Interface

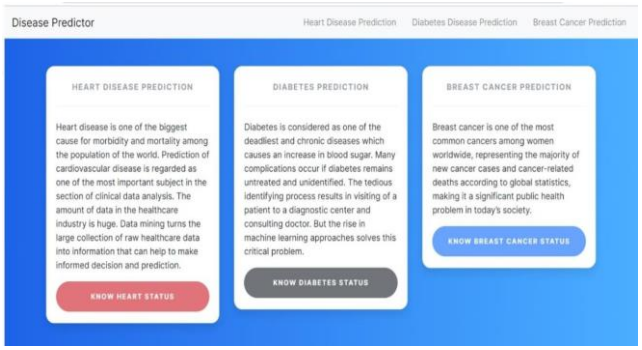

Fig 3: Interface

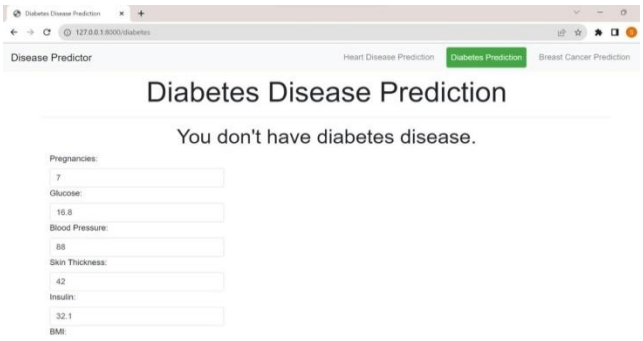

Fig 4: Diabetes disease report

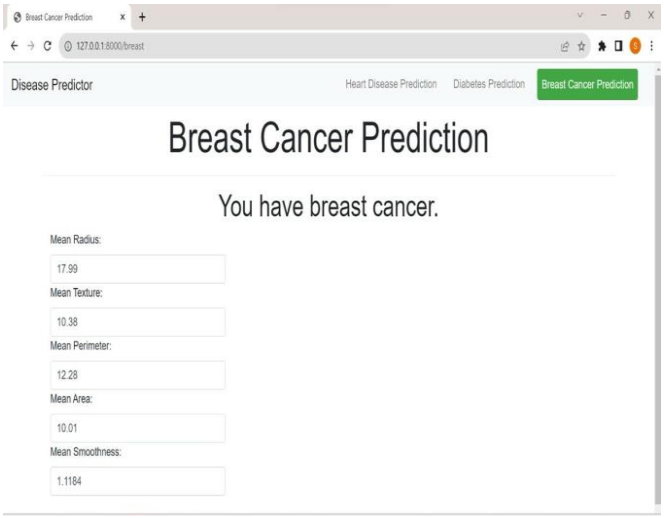

Fig 5: Breast cancer disease report

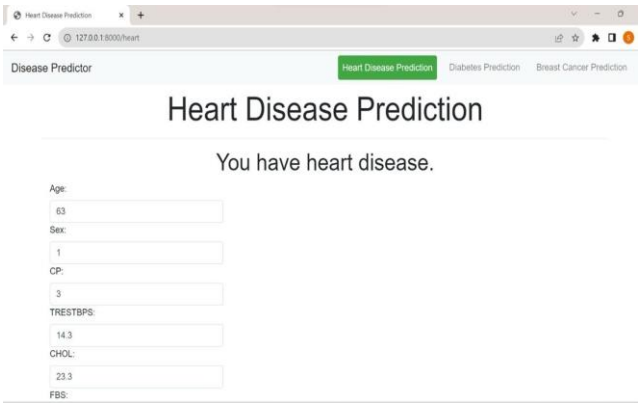

Fig 6: Heart disease report

10. Conclusion

The main goal of the project is to develop a system that can accurately predict various disorders. Thanks to technology, users can save time by avoiding having to switch between multiple websites.

A disease's life expectancy can be increased and financial hardship can be avoided with early diagnosis. We have utilized various machine learning techniques, including Random Forest, to achieve the highest level of accuracy.

- The ability to detect diseases early on will be tremendously helpful to the healthcare sector, especially hospitals.
- We can expand the current system to include more disorders in further.
- To decrease the death tempo, we're able to attempt to go up prediction preciseness.
- Put forth an effort to improve that system's usability.

11. References

- Priyanka Sonar, Prof. K. JayaMalini," DIABETES PREDICTION USING DIFFERENT MACHINE LEARNING APPROACHES", 2019 IEEE ,3rd International Conference on Computing Methodologies and Communication (ICCMC)
- Archana Singh ,Rakesh Kumar, "Heart Disease Prediction Using Machine Learning Algorithms", 2020 IEEE, International Conference on Electrical and Electronics Engineering (ICE3)
- A. Sivasangari, Baddigam Jaya Krishna Reddy, Annamareddy Kiran, P. vAjitha," Diagnosis of Liver Disease using Machine Learning Models" 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC).
- Rajesh. Ranjan, "Predictions for COVID-19 outbreak in India using Epidemiological models", 2020.
- Mohan Senthilkumar, Chandrasegar Thirumalai and Gautam Srivastava, "Effective heart disease prediction using hybrid machine learning techniques", IEEE Access, vol. 7, pp. 81542-81554, 2019.
- Mohan Senthilkumar, Chandrasegar Thirumalai and Gautam Srivastava, "Effective heart disease prediction using hybrid machine learning techniques", IEEE Access, vol. 7, pp. 81542-81554, 2019.
- M. Bayati, S. Bhaskar and A. Montanari, "Statistical analysis of a low cost method for multiple disease prediction", Statistical Methods Med. Res., vol. 27, no. 8, pp. 2312-2328, 2018.