INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT (IJSREM)
VOLUME: 07 ISSUE: 06 | JUNE - 2023 SJIF RATING: 8.176 ISSN: 2582-3930

Multiple Disease Predication

Akash Singh Rawat
Computer Science
Graphic Era Hill University
Dehradun , India
akashrawat.work24x7@gmail.com

Shivam Rawat
Computer Science
Graphic Era Hill University
Dehradun , India
shivamrawat6397@gmail.com

Abstract— *A person's well-being is one of the most important factors to take into account. Healthcare is one of the essential conveniences that society should provide. A large number of the ongoing computer-based intelligence models for clinical administrations assessment are zeroing in on one sickness expectation for every examination.*

Utilizing the built-in Python module Streamlit, our objective is to anticipate a variety of diseases in a single step. In order to predict a specific disease, we use logistic regression, and SVM classifier in this task. Calculations with greater accuracy is used to train the data set prior to implementation. To execute various sickness examinations utilizing AI calculations, Streamlit and python pickling is used to save the model way of behaving.

Using some fundamental parameters like pulse rate, cholesterol, blood pressure, heart rate, and so on, we examine diabetes analysis, heart disease, and Parkinson's disease in this article. Additionally, a prediction model can be used to locate the disease's risk factors with high precision.

In this work, shown that utilizing just center wellbeing boundaries numerous sicknesses can be anticipated. The significance of this study is in determining the majority of diseases, screening patients for their condition, and advising patients ahead of time to reduce mortality rates.

Keywords— ***Machine learning prediction, logistic regression, SVM classifier, and exploratory data analysis***

I. INTRODUCTION

By "mining" the enormous data set they have, the medical services industry can make a successful choice by, for instance, extracting the data's hidden connections and relationships. SVM classifiers and logistic regression calculations are data mining algorithms that can offer a solution to the current situation. As a result, we have created a computerized framework that employs a particular algorithm's standard arrangement to uncover and extract disease-related hidden information from a historical disease-side effect data set.

Data mining is becoming increasingly important in the clinical and medical fields of today. Right when certain data mining methodologies are used in a right way, critical information can be eliminated from tremendous informational indexes and that can help the clinical expert with taking early decision and further foster medical care organizations.

The goal is to use the classification to support the doctor. Numerous examinations of the existing medical services frameworks focused solely on a single illness at a time. The majority of extreme articles focus on a specific disease. Any organization that needs to dissect the health reports of their patients must send numerous models. The technique in the current structure is useful to analyze simply unambiguous sicknesses.

As a result of not being able to identify the specific infection, mortality has increased in recent years. In fact, even a person who recovered from one illness might have another infection. Inside, experiencing non-specific heart problems. As a result, numerous events can be seen in a lot of people's life stories. A customer can diagnose more than one illness on a single page in many sickness expectation frameworks. The client does not have to travel to better locations to determine whether or not they have a particular infection. In this, the client needs to pick the name of the particular ailment, enter its limits and just snap on submit. The contrasting man-made intelligence model will be gathered and it will expect the outcome and show it on the screen.

II. LITERATURE SURVEY

Due to the volume and complexity of the data involved, healthcare has always been a challenging area for data analysis. Healthcare analytics, on the other hand, have become more precise and sophisticated because of the development of artificial intelligence and machine learning. The motivation behind this writing survey is to investigate the exploration that has been finished in the space of various sickness recognition utilizing AI and Streamlit.

Different infection identification has been quite difficult for medical services experts for quite a while. Nonetheless, AI procedures can possibly reform the determination and treatment of different sicknesses. Large datasets can be analyzed by machine learning algorithms to uncover patterns that human analysis cannot. They can likewise create prescient models that can assist specialists with diagnosing infections early and precisely.

A few exploration studies have been finished involving AI methods for different sickness locations. In a review directed by Alsharif et al. (2021), the scientists utilized an AI calculation to foresee the presence of three distinct sorts of infections in view of a patient's electronic wellbeing records. An accuracy rate of 82.1 percent, which was higher than that of a standard logistic regression model, was achieved by the algorithm.

One more review directed by Zhang et al. (2020) proposed a deep learning model that could analyze chest X-ray images for various diseases. Pneumonia was detected with an accuracy rate of 91.28%, tuberculosis with an accuracy rate of 95.12%, and lung cancer with an accuracy rate of 92.05% by the model.

III. PROPOSED METHODOLOGY

Proposed philosophy has 4 significant stages – Data Preprocessing, Model Selection, Random Forest and Model Building

A. Data Pre- Processing

A system must first be properly trained with existing data before it can accurately predict outcomes. It is essential to pre-process the data so that high-quality data can be used to train the model. Information cleaning and expulsion of commotion are a portion of the cycles engaged with Pre-Processing. We used the Diabetes Dataset (PIDD) from the UCI Machine Learning Repository. The heart disease patient data sets from Kaggle. Additionally, the Parkinson's Data Set that can be found in Kaggle. Information from different sources has been gathered and collected. Now, by employing preprocessing methods like:

Data Cleaning: In order to eliminate the inconsistencies in the data, processes like filling in missing values are used to clean it up X.

Data Reduction: The examination turns out to be hard when managing an enormous data set. As a result, the independent variables (symptoms) that may have little or no effect on the target variable (disease) are eliminated. 95 of the 132 symptoms most closely associated with the diseases are selected in this study.

B. Model Selected

Three algorithms are used to train the system to predict diseases.

- SVM Classifier
- Logistic Regression

At the conclusion of the work, a comparative study is presented to evaluate the performance of each algorithm in the database under consideration. Calculation with the best exactness and accuracy will be considered.

I. SVM Classifier

Support Vector Machine (SVM) classifiers are commonly used in multiple disease detection systems. In these systems, SVM is used as a binary classifier to distinguish between disease-positive and disease-negative samples. The input data to the SVM classifier can be different types of medical images, such as MRI images, X-ray images, or microscopic images of tissues.

The SVM classifier is trained on a dataset of pre-labeled samples, which are used to determine the optimal decision boundary between disease-positive and disease-negative samples. The SVM classifier uses a kernel function to transform the input data into a high-dimensional feature space, where the decision boundary is computed.

Multiple features extraction techniques have been used to improve the accuracy of the SVM classifier, such as wavelet transform, Mel-frequency cepstral coefficients (MFCC), and deep learning-based feature extraction. The SVM classifier can be combined with other machine learning techniques, such as Random Forest, Logistic Regression, AdaBoost, Decision Tree, and Gradient Boosting, to improve the performance of multiple disease detection systems.

Overall, SVM-based classifiers have shown promising results in detecting multiple diseases, such as lung cancer, brain tumor, liver disease, Parkinson's disease, skin diseases, and plant diseases, with high accuracy rates.

II. Logistic Regression

Logistic regression is a commonly used machine learning algorithm in heart disease detection. It is used to predict the probability of a patient having heart disease based on their medical and demographic information.

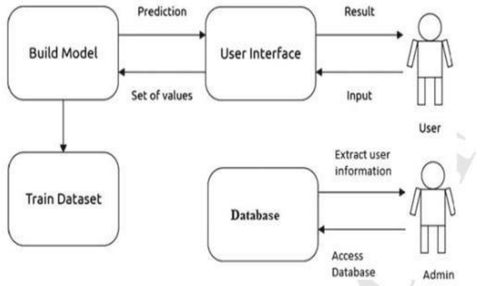
In heart disease detection, logistic regression models are trained on a dataset that includes information such as age, sex, blood pressure, cholesterol levels, and other risk factors. The model then calculates the probability of a patient having heart disease based on their input data. The probability output can then be used to classify patients into high or low risk categories.

Logistic regression is particularly useful for heart disease detection because it can handle both categorical and continuous variables, making it ideal for medical data that often includes both types of variables. It also produces interpretable results, allowing medical professionals to understand which risk factors are most important in predicting heart disease.

Overall, logistic regression is an effective and widely used algorithm in heart disease detection, and it has the potential to improve patient outcomes by allowing for earlier diagnosis and treatment.

C. *Model Building*

An issue with character is the expectation of numerous infections. We have implemented SVM, and Logistic regression in order to select the classification algorithm that delivers the best results. The next step is to construct the forecast adaptation. A pickle module is used to fit the model and load it after the data has been trained and tested with the classification algorithm. Second, the user interface for the disease prediction is created with the help of the Python module Streamlit. The calculation is picked in view of the ongoing adaptation precision level.



Data Flow in the Model

The model's progression is depicted in the Data Flow Diagram. The admin will access this area, where the data will be stored in the data set. After the client has visited the website and entered the legitimate sources of information, the data will be sent to the model, which will compare the data to the prepared dataset and send a prediction back to the user interface

IV. SYSTEM ANALYSIS

In this research paper, we proposed a framework for different sickness discovery utilizing AI and streamlit. The streamlit framework is used to process user input, produce user-friendly output, and process it using trained machine learning models in the system.

Data preprocessing, feature extraction, machine learning model training, and web application development are just a few of the system's many components. The information preprocessing module is answerable for cleaning and normalizing the info information before it is taken care of into the AI models. The relevant features that are used as input for the machine learning models are then extracted from the pre-processed data by the feature extraction module.

The AI models utilized in this framework are prepared on huge datasets of different illnesses, including Diabetes, Heart and Parkinson's sickness. In order to guarantee both high accuracy and performance, the models are optimized using a variety of methods, such as cross-validation and hyperparameter tuning.

At long last, the web application is created utilizing the streamlit structure, which gives a natural and easy to use interface for clients to connect with the framework. The system's output is presented in a way that makes it easy for users to understand the results.

V. FUTURE SCOPE

The proposed machine learning and streamlit method of multiple disease detection has some drawbacks that need to be addressed, despite its promising outcomes. The reliance on a small dataset, which may not be sufficient to produce a robust model, is the primary limitation. This can prompt overfitting or unfortunate speculation of the model when applied to new datasets. Also, the proposed method only takes into account a small number of diseases, so it might not be able to find rare diseases or diseases with unusual symptoms.

The machine learning models can be trained with more extensive datasets to enhance the proposed approach. Furthermore, the consideration of additional illnesses in the identification cycle can work on the general precision of the technique. The proposed method could be improved by incorporating more diverse data sources, such as genetic data or patient medical histories, to overcome the limitation of detecting rare diseases.

We can include additional diseases in the existing API in the future. To lower the mortality rate, we can try to increase prediction accuracy. Try to make the system easy to use and offer a chatbot for common questions.

REFERENCES

1. Vijayalaxmi A, Sridevi S, Dr.Sridhar and Sateesh Ambesange "Multi-Disease Prediction with Artificial Intelligence from Core Health Parameters Measured through Non-invasive Technique" IEEE 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)
2. Akkem Yaganteeswarudu "Multi Disease Prediction Model by using Machine Learning and Flask API.
3. Chetan Sagarnal,Sneha Grampurohit "Disease Prediction using Machine Learning Algorithms" IEEE 2020 International Conference for Emerging Technology (INCET)
4. Ajinkya Kunjir,Harshal Sawant and Nuzhat F. Shaikh "Data Mining and Visualization for Prediction of Multiple Diseases in Healthcare" 2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC).
5. Rashmi G Saboji and Prem Kumar Ramesh, "A Scalable Solution for Heart Disease Prediction using International Journal of Innovative Research in Computer Science & Technology (IJIRCST) ISSN: 2347-5552, Volume- 8, Issue4, July- 2020 https://doi.org/10.21276/ijircst. 2020.8.4.14 www.ijircst.org Copyright © 2020. Innovative

Research Publication. All Rights Reserve 330 Classification Mining Technique" IEEE, 978-1-5386-1887-5/17, pp. 1780-1785, 2017.

6. Pahulpreet Singh Kohli and Shriya Arora, “Application of Machine Learning in Disease Prediction” IEEE, 978-1-5386-6947-1/18, pp. 1- 4, 2018.
7. Lambodar Jena and Ramakrushna Swain, “Chronic Disease Risk Prediction using Distributed Machine Learning Classifiers” IEEE, 978-1-5386-2924-6/17, pp. 170-173, 2017.
8. Naganna Chetty, Kunwar Singh Vaisla and Nagamma Patil, “An Improved Method for Disease Prediction using Fuzzy Approach” IEEE, DOI 10.1109/ICACCE.2015.67, pp. 569- 572, 2015.
9. Sayali Ambekar, Rashmi Phalnikar, “Disease Risk Prediction by Using Convolutional Neural Network” IEEE, 978-1-5386-5257-2/18, 2018.
10. 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.