IMPLEMENTATION OF DISEASE PREDICTION SYSTEM USING MACHINE LEARNING ALGORITHM AND FLASK API(HEART,BREAST CANCER,DIABETES,KIDNEY,LIVER)

S.Vanarasan M.E.,[1],A.Harish[2],Y.Imran Khan[3],S.MadhanKumar[4],V.Tamilarasan[5]

#1 Assist.Professor, Dept.of IT , PSV College of Engineering and Technology,Krishanagiri,Anna University, Tamilnadu,India

#2UGScholar,Dept.of IT,PSV College of Engineering and Technology,Krishanagiri,Anna University, Tamilnadu State,India

#3UGScholar,Dept.of IT,PSV College of Engineering and Technology,Krishanagiri,Anna University, Tamilnadu State,India

#4UGScholar,Dept.of IT,PSV College of Engineering and Technology,Krishanagiri,Anna University, Tamilnadu State,India

#5UGScholar,Dept.of IT,PSV College of Engineering and Technology,Krishanagiri,Anna University, Tamilnadu State,India

Abstract: One disease at a time is the focus of many of the machine learning models for healthcare analysis currently in use. For example, one analysis is done for cancer, another for heart disease, etc. Multiple diseases cannot be predicted using the same framework by a single analysis. This study proposes a system to forecast various diseases using the Flask Application programming interface (API). The approach for examining diabetes, heart disease, kidney, liver, and cancer was proposed in this work. Numerous disease analyses were implemented using TensorFlow, the Flask API, and machine learning techniques. Python dataset is used to save the model behavior and to load the dataset file as necessary. The importance of this article analysis lies in the fact that all the contributing elements to the illness are taken into account while analyzing it, allowing for the identification of the disease's complete range of potential effects. For example, a lot of the existing techniques for analyzing diabetes took into account a few factors like age, sex, BMI, and insulin. However, in contrast to the existing model, our proposed system takes into account additional variables like the number of pregnancies, blood glucose levels, skin thickness, heart rate, and family history of diabetes. The behavior of the finished model will be captured in a Python dataset file, and the Flask API will be developed. When using this API, the disease's parameters must be sent along with the disease's name. The Flask API will call the pertinent model and then return the patient's status. The goal of this project is to look at as many diseases as we can in order to monitor patients' status and provide them warnings, when necessary, in order to lower the death rate

Keywords: Machine Learning, Flask API, Convolutional Neural Network, XG-Boost

1.INTRODUCTION

Artificial intelligence (AI) technologies, which are widely used in modern business and daily life, are progressively being incorporated into healthcare. Healthcare providers could benefit from artificial intelligence in many aspects of patient care and operational processes, allowing them to build on existing solutions and find answers to issues more quickly. Although the majority of AI and healthcare technologies are highly relevant to the healthcare industry, hospitals and other healthcare organizations may use very different strategies. Precision medicine is the most prevalent application of traditional machine learning in the field of artificial intelligence in healthcare. It is a significant advancement for many healthcare organizations to be able to predict which treatment techniques will be most effective with patients based on their features and the treatment framework. And, while some articles on the use of artificial intelligence in healthcare suggest that it can perform as well as or better than humans at certain procedures, such as disease diagnosis, it will be some time before AI in healthcare replaces humans for a wide range of medical tasks. The majority of AI in healthcare is made up of machine learning and precision medicine applications, which require data for training with known results. This is referred to as supervised learning.

Many analyses of the current health-care systems only looked at one ailment at a time. For example, publications are used to study diabetes, brain tumors, and predict heart disease. The majority of articles are focused on a single condition. Any institution that wishes to analyze its patients' health reports must employ a variety of models. The strategy utilized in the current system is useful

for examining only one disease. The failure to effectively diagnose diseases has increased mortality today. Even after receiving treatment for one condition, a patient may develop another. Several current methods used a limited number of parameters to analyze the disease. As a result, it may be impossible to predict which diseases will be caused by the disease's effects. Diabetes, for example, raises the risk of dementia, hearing loss, retinopathy, neuropathy, and heart disease. This research considered the examination of diabetes, heart disease, and kidney, liver, cancer data sets. Several other diseases, such as skin conditions, fever-related diseases, and others, may be added in the future.

When adding a new disease analysis to this API, the developer must include the model file associated with the analysis. When creating new diseases, the developer must configure Python picking to save model behavior. While using this Flask API, the developer can load a pickled file to obtain model behavior. When a user wants to analyze a patient's health, they can anticipate a specific disease or, if the report contains factors that indicate other diseases, this analysis will produce the most appropriate disease identifications. This article tries to reduce the mortality ratio by forewarning patients based on their health circumstances. Because several disease models and forecasts are made in one location, the cost of patient analysis can be reduced

Background study

[1] Priyanka Sonar, Prof. K. Jaya alini," DIABETES PREDICTION USING DIFFERENT MACHINE LEARNING APPROACHES", 2019 IEEE ,3rd International Conference on Computing Methodologies and Communication (ICCMC)

According to the publication, diabetes is one of the most hazardous diseases in the world, causing a wide range of ailments such as blindness. In this paper, scientists employed machine learning approaches to detect diabetic disease since it is simple and adaptable to predict whether or not a patient is ill. The goal of this investigation was to create a system that can help the patient diagnose diabetic disease with accurate results. They compared the accuracy of four main algorithms, Decision Tree, Nave Bayes, and SVM, which are 85%, 77%, and 77.3%, respectively. They also employed an ANN algorithm after the training phase to observe the network's reactions, which indicated whether the sickness was correctly identified or not.

[2] Archana Singh, Rakesh Kumar, "Heart Disease Prediction Using Machine Learning Algorithms", 2020 IEEE, International Conference on Electrical and Electronics Engineering (ICE3)

The major goal of the paper is, because the heart is crucial in living creatures. As a result, the diagnosis and prediction of heart illness should be accurate and correct because it is very important and might cause death cases related to the heart. As a result, machine learning and artificial intelligence can help predict any natural event. So, using the UCI repositor dataset for training and testing, they calculate the accuracy of machine learning for predicting heart disease using k-nearest neighbour, decision tree, linear regression, and SVM. They also compared algorithm accuracy: SVM 83%, Decision tree 79%, Linear regression 78%, and k-nearest neighbour 87%.

[3] A.Sivasangari, Baddigam Jaya Krishna Reddy, Annamareddy Kiran, P.Ajitha," Diagnosis of Liver Disease using Machine Learning Models" 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)

According to the system, liver illnesses are a leading cause of death in India and are also a leading cause of death worldwide. Because early detection of liver disease is difficult. As a result, we can accurately detect liver illness using an automated programme that employs machine learning techniques. For quantitative measurement, they used and compared SVM, Decision Tree, and Random Forest algorithms, as well as precision, accuracy, and recall metrics. The accuracy is 95%, 87%, and 92%, respectively.

2. Existing System

Many of the existing machine learning models for health care analysis focus on a single disease at a time. For example, the first is for liver analysis, the second for cancer analysis, and so on for lung disorders. If a person wishes to anticipate more than one sickness, he or she must visit many websites. There is no standard system in which a single analysis may forecast more than one disease. Some of the models have low accuracy, which can have serious consequences for patients' health. When an organization wishes to examine their patients' health reports, they must deploy numerous models, which increases both the cost and the time required. Some existing systems consider only a few parameters, which can lead to incorrect results.

3. Proposed System

It is feasible to predict more than one disease at a time in multiple disease prediction. As a result, the visitor does not need to visit multiple websites in order to predict diseases. We are focusing on three diseases: liver, diabetes, and heart. Because all three diseases are linked to one another. We will utilise machine learning methods and Flask to implement numerous disease analyses. When accessing this API, the user must send the disease parameters as well as the disease name. Flask will run the associated model and return the patient's state.
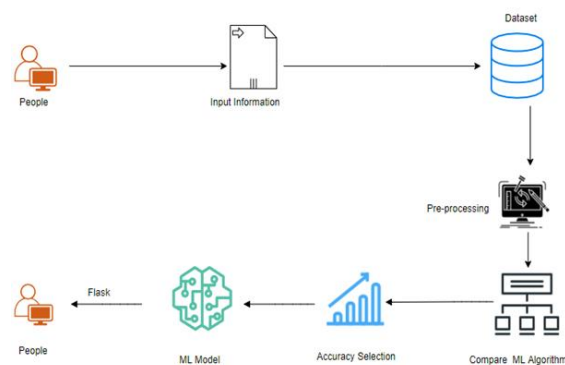
4 .System Design

Fig.1.Architecture diagram

For high-level requirement analysis of a system, use case diagrams are explored. As a result, when a system's needs are analyzed, the functionality are captured in use cases. As a result, it is possible to claim that use cases are nothing more than system functionalities written in an organized manner. When we feed information into our model, it will preprocess it and select the best accuracy model to output to the user. If the model predicts a result of one, the user is more likely to be affected by disease; otherwise, the user is less likely to be affected by disease.
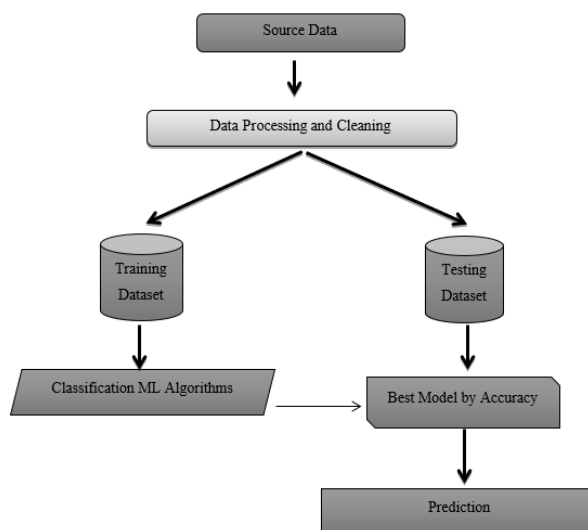

Fig.2 Workflow diagram

5. Implementation

5.1.Algorithm And Techniques
5.1.1 Knn Algorithm
The following is how the K-NN algorithm works:

Step 1: Begin by determining the K value, for example, k=5.

Step 2: Next, we'll calculate the Euclidean distance between the spots. It is calculated as follows:

$$Euclidean\ Distance = \sqrt{(X2 - X1)^2 + (Y2 - Y1)^2}$$

Step 3: Next, we'll compute the nearest neighbour's Euclidean distance.

Step 4: Count the number of data points in each category. For example, three values were discovered for Category A and two values were discovered for Category B.

Step 5: Finally, assign the new point to the category with the greatest number of neighbours. For example, because Category A has the most neighbours, we will allocate the new data point to it.

Step 6: Our Knn model is now complete.

5.1.2 Random Forest Algorithm

A random forest can be created by merging N decision trees, and it can then be used to make predictions for each tree created in the initial phase.

The random forest works like this:

Step 1: It will select K data points at random from the training set.

Step 2: After selecting k data points, construct decision trees that are related to the selected data points (Subsets).

Step 3: Select the Nth node for the decision trees you want to build.

Step 4: In step 4, repeat steps 1 and 2.

Step 5 is finding the predictions of each decision tree and distributing the new data points to the category with the most support.

5.1.3CNN Algorithm

A **Convolutional Neural Network (CNN)** is a type of Deep Learning neural network architecture commonly used in Computer Vision. Computer vision is a field of Artificial Intelligence that enables a computer to understand and interpret the image or visual data.

The working of CNN algorithm

Step 1: The dataset is first converted to vector format.

Step 2: Next, word embedding was finished, with zero values used to fill in the data. As a result of word embedding, a convolutional layer is formed.

Step 3: After passing the convolutional layer as an input to the pooling layer, we perform the maximum pooling operation on it.

Step 4: For Max pooling, convert the dataset into a fixed-length vector form. The pooling layer is connected to the full connected neural network.

Step 5: The SoftMax classifier is connected to the entire connection layer.

5.1.4XG-Boost Algorithm

The working of XGBoost algorithm are as follows:

Step 1:The first step is to make a single leaf tree.

Step 2: After calculating the target variable's average as a forecast for the first tree, we must compute the residuals using the provided loss function. The residuals for future trees are then calculated using the first tree's prediction.

Step 3: Applying the formula to calculate the similarity score.

$$Similarity\ Score = Gradient\ \frac{Gradient^2}{Hessian + \lambda}$$

where, Hessian is equal to number of residuals; Gradient2 = squared sum of residuals; λ is a regularization hyperparameter.

Step 4: We select the relevant node based on the similarity score. When the similarity score is higher, there is more homogeneity.

Step 5: The fifth step entails applying the similarity score to the received information. The amount of homogeneity acquired by dividing the node at a certain point is revealed via information gain, which aids in distinguishing between old and new similarities.

Information Gain = LS + RS – Similarity for Roots

Where LS= Left Similarity & RS=Right Similarity

Step 6: Using the technique, you may prune and regularize the tree to the appropriate length by adjusting the regularization hyperparameter.

Step 7: Using the Decision Tree you created, we can then forecast the residual values.

Step 8: The learning rate is used to calculate the new set of residuals.

Step 9: Next, go back to step 1 and carry out step 1 again for each tree.

6. Conclusion

We employed typical machine learning algorithms to categorise patient data because medical data is rapidly expanding in the modern medical world and must be processed to produce exact disease predictions based on symptoms.We were able to develop an accurate general illness risk prediction by supplying the input of patient records, which aid in determining the degree of disease risk prediction. This technique allows for illness and risk prediction with minimal effort and money. In terms of accuracy and processing time, we compare the results of the various algorithms. The CNN method outperforms the other machine learning algorithms in terms of accuracy, and its processing time is shorter. As a result, CNN outperforms other machine learning algorithms in terms of accuracy and timing.

7.Appendix

Fig.3 Flask API

Fig.4 DataSet

Fig.5 Prediction page

8.Acknowledgements

REFERENCES

[1] Priyanka Sonar, Prof. K. JayaMalini," DIABETES PREDICTION USING DIFFERENT MACHINE LEARNING APPROACHES", 2019 IEEE ,3rd International Conference on Computing Methodologies and Communication (ICCMC)

[2] Archana Singh ,Rakesh Kumar, "Heart Disease Prediction Using Machine Learning Algorithms", 2020 IEEE, International Conference on Electrical and Electronics Engineering (ICE3)

[3] A.Sivasangari, Baddigam Jaya Krishna Reddy,Annamareddy Kiran, P.Ajitha," Diagnosis of Liver Disease using Machine Learning Models" 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)