Smart Disease Predictor Using Machine Learning

[1] **Navaneetha Krishnan M,** [2] **Prasanth G,** [3] **Gowtham R**

[1] **Assistant Professor,** [2, 3] **Scholar,** [1, 2 ,3] **Department of MCA,** [1, 2] **Karpagam College of Engineering, Coimbatore, India.** [3] **PSG College of Arts and Science, Coimbatore, India**

[1] **navaneethanmca@gmail.com,** [2] **prasanth287471@gmail.com,** [3] **gowthamr0205@gmail.com**

Abstract - **Most of the present machine studying sickness prediction models for health care prediction focus on a single illness per analysis. One analysis may be for heart illness, another for cancer, and still another for diabetic problems. There's no single system in which one evaluation can carry out multiple disease illness prediction. This article proposes a system that uses machine learning and the Flask API to forecast numerous illnesses. Kidney ailment, liver illness, heart ailment, and breast cancer, diabetes illness are all mentioned in this article, later other diseases like Malaria, Pneumonia evaluation and lots of greater sicknesses disease prediction will be included in future. Machine learning techniques, TensorFlow, and the Flask API were utilized to accomplish various illness prediction. Python pickling is used to keep the model behavior and python unpickling is used to load the pickle file whenever required. The significance of this article analysis is that it includes all of the disease-causing parameters, making it feasible to detect the disease's maximal impacts. A Heart Disease Dataset, for example, was trained using various machine learning algorithms in this research. Predictions were made using the algorithm with the best accuracy percentage. The final model will be stored as a pickle file in Python. In order to predict result, the model uses various input from the user. With using Flask API, the user must supply the disease's parameters as well as the disease's name to the system. Flask API will call the associated model and return the patient's status. The accuracy percent of the prediction results is higher, which is quite encouraging. The significance of this evaluation is to examine the most diseases, in order that to screen the patient's circumstance and warn the sufferers earlier to lower mortality ratio.**

Keywords — Diseases, Prediction, Flask API, TensorFlow, Dataset

I. INTRODUCTION

Many studies of existing health-care analytic prediction systems focused on only one disease at a time. Article [1] is used to analyses diabetes illness, article [2] is used to breast cancer disorder, and article [3] is used to forecast heart disease, for instance. The present system's technique is suitable for analyzing only a single disease at a time.

The current increase in death is related to the failure to identify the exact sickness. Even if a patient has been recovered of one disease, they may still be suffering from another. It was my experience in real life. My uncle was recovered of his liver disease and was discharged from hospital, but he died a few days later. Internally, he has unknown cardiac problems. Several cases like these were seen in many people's life stories.

When analyzing the illness, several current systems evaluated only a few factors. As a result, it may be difficult to determine the precise percentage of diseases that will be induced by the disease's influence for example, Diabetes, can lead to heart disease, blindness, loss of hearing, and insanity etc. The data sets for Kidney Disease, Liver Disease, Heart Disease, Diabetes and Breast Cancer analyses were utilized to develop the model in this article. Many additional diseases, such as Malaria, Pneumonia, and fever-related disorders, will be included in the future. This versatile analysis eventually includes various disorders for evaluation. The developer must provide the model file for the new diseases analysis when adding any new disease analysis to this present API.

When creating a new illness prediction, the developer must prepare Python pickling in order to save model behavior. The developer may use this Flask API to load a pickled file to extract the model behavior that is used to forecast illnesses. The major goal of this article is to avoid the death percentage from growing day by day by advance notice patients depending on their health status. Because multiple illnesses models and forecasts are done in one spot, the cost of patient analysis may be minimized.

II. PROPOSED WORK AND MODEL DESIGN

A. Existing System

Many of the existing studies focused on a one particular disease. When a user wants to analyses diabetes [1], they must use another model and when they want to analyze heart disease [3], they must use another model. This is a very hard procedure. Also, if a user has different diseases but the present system can only predict one of diseases, the death rate may rise as a result of the inability to anticipate the other sickness in advance.

B. Proposed System

This prediction model would forecast more than one disease at the same time. As a result, the user does not have to go through several models in order to predict illnesses. It will save time, and it has the ability to lower death rates by detecting numerous disorders in one go.

C. Dataset Preparation

Multiple illness datasets have been collected using real-time patient data from various machine learning dataset resources [15], including diabetes, heart disease, and breast cancer, liver disease, and kidney disease. The significance of the study is that it collects the key parameter that is used to predict illness. Because of this study, the morality ratio will be reduced, allowing doctors to advise patients about the procedure ahead of time.

D. Machine Learning Techniques

Article main objective is to build a multi disease prediction model so machine learning techniques used are briefly summarized here. Diabetes analysis, heart disease prediction and Breast cancer detection, Liver Disease and Kidney disease are analyzed by different machine learning techniques. Like Naïve Bayes [12], K-Nearest Algorithm [10], Decision Tree Algorithm, and Random Forest Algorithm [9] which will help us in getting accurate predictions. Our aim with this system is to be the connecting bridge between doctors and patients. And also, will find which algorithm gives a faster and efficient result.

III. MODEL BEHAVIOUR SAVING WITH PYTHON PICKLING

A. Python Pickling for Breast Cancer Disease Prediction

Python Pickling for Example Breast Cancer Disease Prediction Model After processing the data using the data sets, the appropriate algorithm with the highest accuracy is chosen and used algorithms are random forest, Decision tree etc.

Python pickling [13] can be used to store model results. The pickle mechanism in Python is used to pre-process and counter Python object structures. Python objects can be pickled and stored in memory. The python pickle file is a character stream that contains all of the variable information required to replicate the item in another module.

Consider hear.csv is a Breast Cancer disease csv file. To process the file and saving it as a pickle file [13], see the below code

For Example: -

Breast Cancer disease model file pickling:

```
from sklearn.model_selection
import train_test_split
from sklearn.ensemble import RandomForestClassifier
model = RandomForestClassifier(n_estimators=20)
model.fit(X_train, y_train)
classifier.fit(X_train, y_train)
import pickle
pickle.dump(classifier, open('breast_cancer.pkl', 'wb'))
```

In the above sample code 'breast_cancer.pkl', the pickle file is created. As like as breast cancer prediction pickle file other disease pickle files also will be created.

B. Python Pickling for Kidney Disease Prediction

Kidney disease model file pickling:

```
from sklearn.metrics
import confusion_matrix, accuracy_score
confusion_matrix(y_test, model.predict(X_test))
print(f"Accuracy is {round(accuracy_score(y_test, model.predict(X_test))*100, 2)}%")
import pickle
pickle.dump(model, open('kidney.pkl', 'wb'))
```

C. Python Pickling for Heart Disease Prediction

Heart disease model file pickling:

```
classifier.fit(X_train, y_train)
confusion_matrix(y_test, classifier.predict(X_test))
print(f"Accuracy is {round(accuracy_score(y_test, classifier.predict(X_test))*100,2)}%")
import pickle
pickle.dump(classifier, open('heart.pkl', 'wb'))
```

D. Python Pickling for Diabetes Disease Prediction

Diabetes disease model file pickling:

   model.fit(X_train, y_train)

   from sklearn.metrics import accuracy_score

   print(accuracy_score(y_test,
   model.predict(X_test))*100)

   import pickle

   pickle.dump(model, open("diabetes.pkl",'wb'))

E. Python Pickling for Liver Disease Prediction

 Liver disease model file pickling:

   model.fit(X_train, y_train)

   from sklearn.metrics import confusion_matrix,
   accuracy_scoreconfusion_matrix(y_test,
   model.predict(X_test))

   import pickle

   pickle.dump(model, open('liver.pkl', 'wb'))

IV. DISEASE PREDICTION MODEL

A. Loading Pickle File to Predict the Disease

   Uploading pickle file to predict the disease Once the model building is finished and model behaviour saved as pickle file. In this analysis considered 5 diseases, so 5 pickle files are generated. Before analyzing the disease load the all-pickle files in the python script where the multi diseases analysis takes place.

   model = pickle.load(open('models/diabetes.pkl','rb'))

   model = pickle.load(open('models/breast.pkl','rb'))

   model = pickle.load(open('models/heart.pkl','rb'))

   model = pickle.load(open('models/kidney.pkl','rb'))

   model = pickle.load(open('models/liver.pkl','rb'))

B. Data Flow to Predict the Disease

   Figure 1 represents data flow in analyzing the prediction. The analysis begins with the dataset being pre-processed. Because, for example, a real human being's blood pressure cannot be zero, pre-processing is necessary. Such records should be pre-processed. After pre-processing the dataset, various ailments model building takes place using datasets. Pickle files include all of the model's behaviors. The Prediction Flask API was created with Pickle file [14]. Based on the User input the result will be predicted.
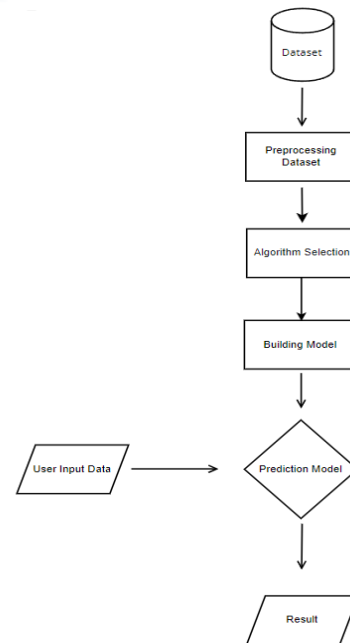

Fig. 1 Data flow of the Smart Disease Predictor using Machine Learning

Sample code for

A. Breast Cancer Disease prediction Flask API:

@app.route("/cancer", methods=['GET', 'POST'])

def cancerPage():

return render_template('breast_cancer.html')

B. Liver Disease Prediction Flask API:

@app.route("/liver", methods=['GET', 'POST'])

def liverPage():

return render_template('liver.html')

C. Heart Disease Prediction Flask API:

@app.route("/heart", methods=['GET', 'POST'])

def heartPage():

 return render_template('heart.html')

D. Kidney Disease Prediction Flask API:

 @app.route("/kidney", methods=['GET', 'POST'])

 def kidneyPage():

 return render_template('kidney.html')

E. Diabetes Disease Prediction Flask API:

 @app.route("/diabetes", methods=['GET', 'POST'])

 def diabetesPage():

return render_template('diabetes.html')

Once Flask API is designed load the pickle file and return the patient status to the user.

V.   RESULT ANALYSIS

Once Flask API is designed. Model can be consumed at front end. Verified by designing sample website Figure 2 shows heart disease input screen. Once user clicks on get status of heart disease patient it will return whether patient have heart disease or not.
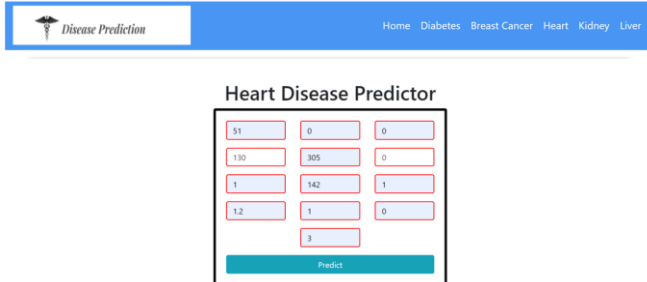

Fig.2. Heat disease prediction input screen

Figure 3 represents output of the heart disease prediction. Once the user enters all details and click on get heart disease status. Corresponding model will be loaded and result of heart disease will be shown to the user.
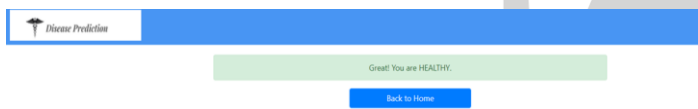
Fig.3.Patient Status Output of Fig2

VI.   CONCLUSION

Multi disease prediction model is used to predict multiple diseases at a time. In this machine learning approach four algorithms were used to train and analyses the dataset which contains the test results of different patients and these algorithms were also tested for accuracy plotting a graph using matplotlib. After testing for accuracy using Decision Tree, Random Forest and K Nearest Neighbors algorithms have the highest accurate result with better patient care. Here based on the user input disease will be predicted. The choice will be given to user. If the user wants to predict particular disease based on user entered inputs corresponding disease model will be invoked and predicted. The advantage of multi disease prediction model in advance can predict the probability of occurrence of various disease and also can reduce mortality ratio.

ACKNOWLEDGMENT

REFERENCES

[1]   Rani, KM. (2020). Diabetes Prediction Using Machine Learning. International Journal of Scientific Research in Computer Science, Engineering and Information Technology. 294-305. 10.32628/CSEIT206463.

[2]   Rawal, Ramik. (2020). Breast Cancer Prediction Using Machine Learning.7.

[3]   Shah, D., Patel, S. & Bharti, S.K. Heart Disease Prediction using Machine Learning Techniques. SN COMPUT.SCI. 1, 345(2020). https://doi.org/10.1007/s42979-020-00365-y

[4]   Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes—2018 American Diabetes Association Diabetes Care 2018; 41(Supplement 1): S13–S27. https://doi.org/10.2337/dc18- S002

[5]   M.M.Mehdy, E.E.Shair and P.Y.Ng, "Artificial Neural Networks in Image Processing for Earlier Detection of Breast Cancer", Hindawi, Computational and Mathematical Methods in Medicine, Volume 2017, Article ID 2610628

[6]   Diabetes Prevention Program Research Group Long-term effects of lifestyle intervention or metformin on diabetes development and microvascular complications over 15-year follow-up: the Diabetes Prevention Program Outcomes Study.Lancet Diabetes Endocrinol. 2015; 3: 866-875

[7]   Thota R.N.Acharya S.H.Garg M.L.Curcumin and/or omega-3 polyunsaturated fatty acids supplementation reduces insulin resistance and blood lipids in individuals with high risk of type 2 diabetes: a randomised controlled trial.Lipids Health Dis. 2019; 18: 31

[8]   Cosentino F, Grant PJ, Aboyans V, Bailey CJ, Ceriello A, Delgado V, et al. ESC Scientific Document Group. 2019 ESC Guidelines on diabetes, pre-diabetes, and cardiovascular diseases developed in collaboration with the EASD. Eur Heart J 202041:255-323. doi: 10.1093/eurheartj/ehz486.

[9]   Akkem, Yaganteeswarudu. (2020). Multi Disease Prediction Model by using Machine Learning and FlaskAPI.

[10]   https://www.analyticsvidhya.com/blog/2021/04/simple-understanding-and-implementation-of-knn-algorithm/

[11]   https://www.kaggle.com/

[12]   https://www.educba.com/naive-bayes-algorithm/

[13]   https://www.delftstack.com/howto/python/python-read-pickle/

[14]   https://www.geeksforgeeks.org/pickle-python-object-serialization/

[15]   https://www.educba.com/flask-api/