Multi Disease Prediction Model by using Machine Learning

Akkem Yaganteeswarudu,IEEE

Abstract-- **Many of the existing machine learning models for health care analysis are concentrating on one disease per analysis. Like one analysis if for diabetes analysis, one for cancer analysis, one for skin diseases like that. There is no common system where one analysis can perform more than one disease prediction. In this article proposing a system which used to predict multiple diseases by using Flask API. In this article used to analyse Diabetes analysis, Diabetes Retinopathy analysis, Heart disease and breast cancer analysis. Later other diseases like skin diseases, fever analysis and many more diseases can be included. To implement multiple disease analysis used machine learning algorithms, tensorflow and Flask API. Python pickling is used to save the model be haviour and python unpickling is used to load the pickle file whenever required. The importance of this article analysis in while analysing the diseases all the parameters which causes the disease is included so it possible to detect the maximum effects which the disease will cause. For example for diabetes analysis in many existing systems considered few parameters like age, sex, bmi, insulin, glucose, blood pressure, diabetes pedigree function, pregnancies, considered in addition to age, sex, bmi, insulin, glucose, blood pressure, diabetes pedigree function, pregnancies included serum creatinine, potassium, GlasgowComaS cale, heart rate/pulse Rate, respiration rate, body temperature, low density lipoprotein (LDL), high density lipoprotein (HDL), TG (Triglycerides).**

Final models behaviour will be saved as python pickle file. Flask API is designed. When user accessing this API, the user has to send the parameters of the disease along with disease name. Flask API will invoke the corresponding model and returns the status of the patient. The importance of this analysis to analyse the maximum diseases, so that to monitor the patient's condition and warn the patients in advance to decrease mortality ratio.

Keywords-- Flask API, lipoprotein, GlasgowComaScale, Diabetes, Triglycerides, Mortality

I. INTRODUCTION

During a lot of analysis over existing systems in health care analysis considered only one disease at a time. For example, article [1] is used to analyse diabetes, article [2] is used to analyse diabetes retinopathy, article [3] is used to predict heart disease [11]. Maximum articles focus on a particular disease. When any organization wants to analyse their patient's health reports then they have to deploy many models. The approach in the existing system is useful to analyse only particular disease. Now a day's mortality got increased due to exactly not identifying exact disease. Even the patient got cured from one disease may be suffering from another disease. In real life, I faced that situation. My father got cured from the accident. My father got discharged from hospital but after a few days he got expired. Internally suffering from heart issues which is not identified. Like this many instances observed in many people's life stories.

Some existing systems used few parameters while analysing the disease. Due to that may be not possible to identify the diseases which will be caused due to the effect of that disease. For example, due to diabetes, there may be chance of heart disease, neuropath, retinopathy, hearing loss, and dementia.

In this article considered Diabetes analysis, Diabetes Retinopathy, heart disease and cancer detection data sets. In future many other diseases like skin diseases can be included, fever related diseases and many more. This analysis is flexible that later included many diseases for analysis. While adding any new disease analysis to this existing API, the developer has to add the model file related to the analysis of the new disease.

When developing new disease the developer have to prepare python picking to save model behaviour. When using this Flask API, the developer can load pickled file to retrieve the model behaviour. When user wants to analyse the patient's health condition either then can predict a particular disease or if the report contains parameters which are used to predict other diseases then this analysis will produce maximum identification of relevant diseases.

The aim of this article is used to prevent mortality ratio increasing day by day by warning the patients in advance based on their health conditions. Due to many diseases models and predictions done at one place cost of patient analysis can be reduced.

II. PROPOSED WORK AND PROCUDURE FOR MODEL DESIGN

A. Existing system

Many of existing analysis involved analysing particular disease. When a user wants to analyse diabetes needs to use one analysis and same user wants to analyse heart disease then user has to use one more model. This is a time taking process. And also if any user having more than one disease but in existing system if it is able to predict only one disease then there is a chance of mortality rate increase due to not able to predict the other disease in advance.

B. Proposed system

In multi disease model prediction, it is possible to predict more than one disease at a time. So user no need to traverse many models to predict the diseases. It will reduce time and also due to predicting multiple diseases at a time there is a chance of reducing mortality rate.

C. Dataset preparation

For diabetes analysis initially Pima Indian Diabetes Dataset, the data set which was acquired from a hospital in Frankfurt, Germany are used. For diabetic retinopathy over 150 GB image data from the UCI machine learning repository are used. For heart disease analysis Cleveland, Hungarian and Switzerland heart disease patient's data sets are used. And for cancer disease prediction used Breast Cancer Wisconsin (Diagnostic) Data Set which is available in machine learning repository. In the current analysis in addition to those data sets used other live data sets by visiting corresponding hospitals. The importance of this analysis is by consulting the doctors collected the necessary parameters which will cause the disease and also due to that disease any other disease likely to occur. After doing this analysis there is a chance of reducing mortality ratio because if can able to predict the maximum diseases chances of occurring so that can warn the patients in advance for treatment

As per industry standards train set and test are prepared. By using Scikit learn train_test_split method to split the data as 70 % for training and 30 % for testing are divided. Example: diabetes_feature_train, diabetes_feature _test, diabetes_label_train, diabetes_label _test = train_test_split (diabetes_features, diabetes_label, test_size=0.3, random_state=0)

B. Machine learning and Deep learning

Article main focus is to build a multi disease prediction model so machine learning and deep learning techniques used are briefly summarized here. Diabetes analysis, Heart disease prediction and cancer detection are analysed by different machine learning and deep learning techniques. Like logistic regression, Naïve Bayes [13] classification

algorithm, SVM, Decision tree algorithm, Random forest algorithm and many more algorithms are used to find the status of the patient. For diabetes analysis , logistic regression results 92% accuracy, for heart disease classification Randomforest yield 95% accuracy and for cancer detection SVM yield 96 % accuracy.

Diabetes retinopathy analysis contains retina images. So used the python tensorflow library to analyse the images . The tensorflow convolution neural networks is used for building the model and tested with the test set. The built model produced 91 % accuracy.

III. MODEL BEHAVIOUR SAVING WITH PYTHON PICKLING

A. Python pickling for heart disease prediction data set

Once the data set is processed with training set and test set, best algorithm which producing the highest accuracy is selected. Model behaviour can be saved by using python pickling. The python pickle [12] model is used to serialize or de-serializing the python object structure. Python object can be pickled and it can be saved on disk. The python pickle file is a character stream which contains all the necessary information to reconstruct the object in another script.

Consider hear.csv is a heart disease csv file. To process the file and saving it as a pickle file, see the below code

For heart disease model file pickling:

```
from sklearn.ensemble import RandomForestClassifier

import pickle as p

rf = RandomForestClassifier (random_state=x)

rf.fit (X_train, Y_train)

p.dump (rf, open ('final_heartdiseasepred.pickle', 'wb'))
```

In the above sample code 'final_heartdiseasepred.pickle', the pickle file is created.

B. Python pickling for diabetes prediction data set

For diabetes model file pickling:

```
df=pd.read_csv ("diabetes.csv")

x = df.iloc [:, 0:8]

y = df.iloc [:, 8]

x_train, x_test, y_train, y_test = train_test_split(x, y, random_state=0)

sv = svm.SVC (kernel='linear')

sv.fit (x_train, y_train)

p.dump (rf, open ('final_diabetes.pickle', 'wb'))
```

C. Diabetes retinopathy prediction and model saving by tensorflow

Since diabetes retinopathy analysis contains retina images. So tensorflow model is used to analyse the model and model behaviour stored in a file

```
model.compile (loss='categorical_crossentropy',
        optimizer='sgd',
        metrics= ['accuracy'])
```

```
model.fit (X_train, Y_train, epochs=200, validation_data=
(X_valid, Y_valid), verbose=1)
```

```
model.save
('/home/Yaganteeswarudu/diabeticRetino2/my_model.h5')
```

D. Python pickling for cancer prediction data set

```
from sklearn.svm import SVC

svc_model= SVC ()

svc_model.fit (x_train, y_train)

p.dump (svc, open ('final_cancer.pickle', 'wb'))
```

IV. MULTI DISEASE PREDICTION MODEL

A. Loading pickle file to predict the disease

Once the model building is finished and model behaviour saved as pickle file. In this analysis considered 4 diseases, so 4 pickle files are generated. Before analysing the disease load the all pickle files in the python script where the multi diseases analysis takes place

Sample code:

```
heartmodelfile='final_heartdiseasepred.pickle'

diabetesmodelfile='final_diabetes.pickle'

diabetesretinomodelfile='final_cancer.pickle '

diabetereretino='my_model.h5' heartmodel=p.load

(open (heartmodelfile,'rb')) diabetesmodel=p.load

(open (diabetesmodelfile,'rb')) cancermodel=p.load

(open (cancermodelfile, 'rb')) retinomodel =

load_model (diabetereretino)
```

B. Data flow

Figure1 represents data flow in analysis. Analysis starts by pre-processing the data. Pre-processing is required because, for example a live human being blood pressure [11] cannot be zero and so on. Those kind of records to be pre-processed. After pre-processing data set preparation and model building for the different diseases [13] with available data sets. All the model behaviours are saved in pickle files. Flask API designed
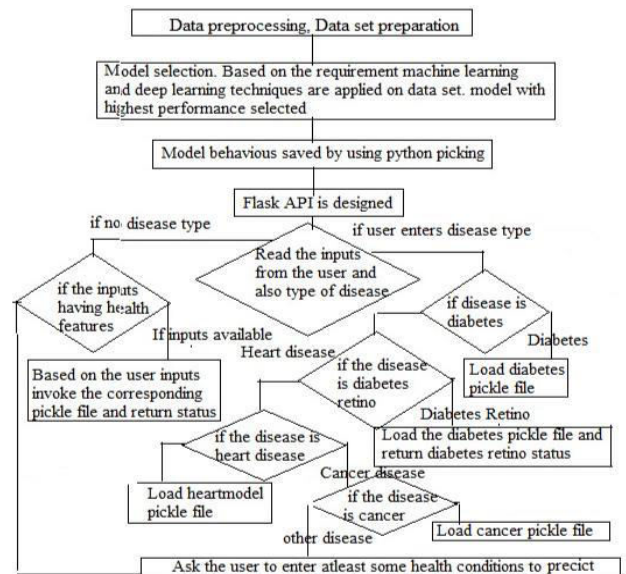

Fig. 1. Data flow

Sample code for Heart Disease prediction Flask API:

```
@app.route ('/heartdiseasecondition', methods= ['POST'])

def heartdiseasecondition():

    url="http://localhost:5000/heartdiseaseprediction"

    # Get the required inputs from user filled form
```

```
data=[[age,sex,cp,trestbps,chol,fbs,restecg,thalach,exang,old
peak,slope,ca,thal]]

    j_data=json.dumps(data)

    headers={'content-type':'application/json','Accept-
Charset':'UTF-8'}

    r=requests.post (url, data=j_data, headers=headers)
```

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

V. RESULTS ANALYISIS

Once Flask API is designed. Model can be consumed at front end. Verified by designing sample website

Figure2 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 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.
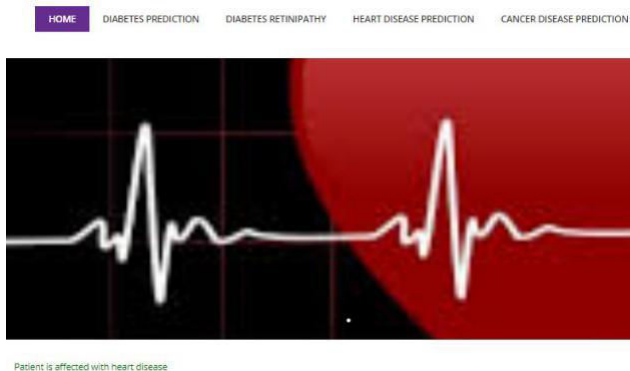

Fig 3: Output of patient status

CONCLUSION

Multi disease prediction model is used to predict multiple diseases at a time. Here based on the user input disease will be predicted. The choice will be given to user. If the user want to predict particular disease or if the user don't enter any disease type then 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.

ACKNOWLEDGM ENT

REFERENCES

[1] 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. IEEE Transactions on health care.

[2] 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

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

[4] DhafarHamed, Jwan K. Alwan, Mohamed Ibrahim, Mohammad B. Naeem "T he Utilisation of Machine Learning Approaches for Med-
ical Data Classification" in Annual Conference on New T rends in Information & Communications T echnology Applications - march-2017 A. J. Jenkins, M. V. Joglekar, A. A. Hardikar, A. C. Keech, D.
N. O'Neal, and A. S. Januszewski, "Biomarkers in diabetic retinopathy," ,IEEE Transactions on multi disesase, vol. 12, no. 1 -2,159–195, 2015.

[5] Jamilian M.et al.T he influences of vitamin D and omega-3 co-supplementation on clinical, metabolic and genetic parameters in women with polycystic ovary syndrome.J Affect Disord. 2018; 238: 32-38"International clinical diabetic retinopathy disease severity scale detailed table", T ech. Rep., 2002.

[6] Chaitrali S. Dangare, Sulabha S. Apte, "Improved Study of Heart Disease Prediction System using Data Mining Classification T echniques", International Journal of Computer Applications (0975 888)Volume 47No.10, June 2012. IEEE Transactions on machine learning.

[7] Alapatt, B.P., Kavitha, A., Amudhavel, J., "A novel encryption algorithm for end to end secured fiber optic communication", (2017) International Journal of Pure and Applied Mathematics, 117 (19 Special Issue), pp. 269-275.

[8] Amudhavel, J., Inbavalli, P., Bhuvaneswari, B., Anandaraj, B., Vengattaraman, T., Premkumar, K., "An effective analysis on har - mony search optimization approaches", (2015) International Journal of Applied Engineering Research, 10 (3), pp. 2035 -2038.

[9] Applying k-Nearest Neighbour in Diagnosing Heart Disease Patients Mai Shouman, Tim T urner, and Rob Stocker International Journal of Information and Education Technology, Vol. 2, No. 3, June 2012

[10] Ponrathi Athilingam, Bradlee Jenkins, Marcia Johansson, Miguel Labrador "A Mobile Health Intervention to Improve Self-Care in Patients With Heart Failure: Pilot Randomized Control T rial" in JMIR Cardio 2017, vol. 1, issue 2, pg no:1 A. Rairikar, V. Kulkarni, V. Sabale, H. Kale, A. Lamgunde, "Heart disease prediction using data mining techniques", 2017 International Conference on Intelligent Computing and Control (I2C2), pp. 1 -8, 2017, June.

[11] T hota 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

[12] 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 2020 41:255 -323. doi: 10.1093/eurheartj/ehz486.

[13] Valensi P.,Prévost G.,Schnell O.,Standl E.,Ceriello A. T argets for blood glucose: what have the trials told us.Eur J Prev Cardiol. 2019; 26: 64-

Authors:

Akkem Yaganteeswarudu, Finished post-graduation (M .Tech) in 2012. Started professional career in 2012.

Worked at Honeywell, ITC InfoTech, Unisys and currently working at Pyramid Soft Sol as a Data scientist. Having 8+ years of experience and in Data science having 4+ years of experience. Published various IEEE papers like preventing suicides of farmers, Speaking compiler and many more. Having interest in research in health care domain. Interested to develop health care solutions by using machine learning and deep learning techniques.