MULTIPLE DISEASE PREDICTION USING MACHINE LEARNING

Vinayak Jadhav, MCA Student, IIMS
Prof. Nitin Ganeshar Assistant Professor, IIMS, Chinchwad, Pune.

Abstract
With in the todays times, people experience a extend of ailments as a result of natural variables and their way of life choices, underscoring the importance of early disease discovery.. So our venture primarily bargains with this major issue. Which includes, when any individual is having any wellbeing issue or affected with some disease or some abnormal things in their body, they utilized to visit a specialist. The issue is that numerous individuals cannot get it the kind of issues thatthey are having base donthe indications . The infections which are predicted in this research are heart disease , kidney disease and diabetes . Our proposed work executes a framework that predicts different illnesses based on the side effects by utilizing machine learning Algorithms like Random Forest , Decision Tree , K-Nearest Neighbour Support Vector Machine .

Keywords—
Random Forest, Decision Tree , K Nearest Neighbour , Support Vector Machine , Diabetes, Heart Disease, Kidney Disease, Breast Cancer.

I. INTRODUCTION

With in the todays times , people experience a extend of ailments as a result of natural variables and their way of life choices , under scoring the importance of early disease discovery.. So our venture primarily bargains with this major issue. which includes, when any individual is having any well being issue or Now a days the different human creatures are influencing with sicknesses and they visit specialists to urge the rapy of the sickness. The issue is that numerous individuals affected with a few diseases.

In this Medical field, data lot of data is generated on regular basis. Data in the Medical filed consists of all the information related to patients. Here architecture has been proposed for predicting the disease. Many of the existing models are predicting the one disease per analysis. Like one analysis for diabetes, one for heart analysis, one for Parkinson's diseases like that. There is no common systemis developed that can analyse more than one disease at a time.Thus, we are concentrating on providing system where patent can do the accurate disease predictions according to the symptoms they enter.

Disease prediction can recognize patientsat chance of disease or health conditions.Due to the later progression of devices and methods for information analytics, disease chance prediction can use expansive sums of semantic data, such as socioeconomics, clinical determination and measurements,health practices, research facility comes about, prescriptions and care utilization. In this respect, electronic health information can be a potential choice for developing disease prediction models. Here we are mainly focusing on a few of the diseases like diabetes, heart disease, kidney disease, Parkinson's disease etc.

II. METHODOLOGY

The methodology for the Multiple Disease Prediction can be summarized as follows:

1. Data Collection :Data is collected from Kaggle.com, a popular platform for accessing datasets. The data isobtained specifically for diabetes, heart disease, kidney disease, Parkinson's disease.
2. Data Preprocessing: The collected data undergoes preprocessing to ensure its quality and suitability for training the DATA.
3. Model Selection: Different machinelearning algorithms arechosen for each diseaseprediction task. Support Vector Machine (SVM), Logistic Regression, and TensorFlow with Keras are selected as the algorithms for various diseases based on their performance and suitability for the specific prediction tasks.
4. Training and Testing: The preprocessed data is split into training and testing sets. The models are

trained using the training data, and their performance is evaluated using the testing data. Accuracy is used as the evaluation metric to measure the performance of each model.
5. Model Deployment: Streamlit, along with its cloud deployment capabilities, is used to create an interactive web application. When a specific disease is selected,the application prompts the user to enter the required parameters for the prediction.

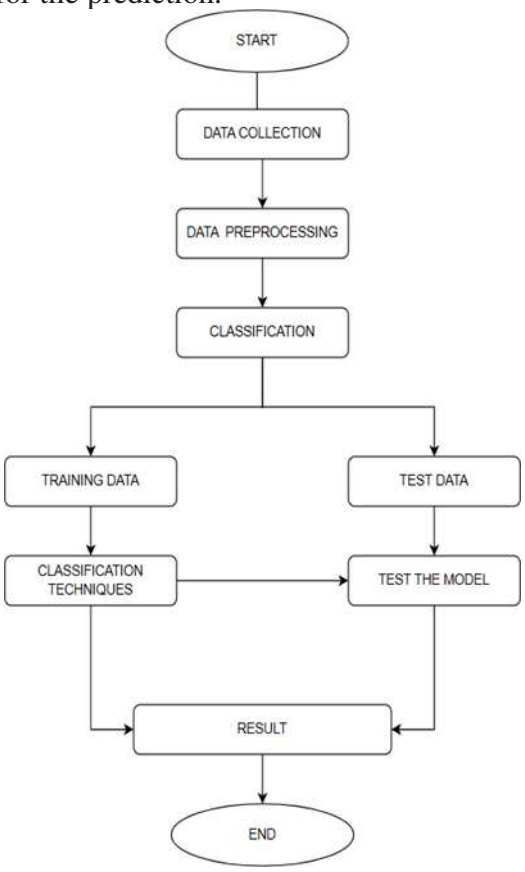

Fig.1:Methodology of Multiple Disease Prediction

III. BACKGROUND

With the use of machine learning , the categorization that is currently in place was farbetter . In the sphere of medicine
, these algorithms were helpful in making predictions . They are even capable of being trained on any kind of data using that specific model . And inorder to do so ,there area number of information cleaning techniques accessible that are helping in the expulsion of undesirable information from the dataset. The best challenge was utilizing themachine learning calculations themselves to estimate the information and expel the lost values from the dataset.

IV. REVIEW OF LITERATURE

According to the text, this article will focus on diabetes as a dangerous disease in the world. Diabetes can cause many conditions, including blindness. In this article, the author uses machine learning algorithms to detect diabetes. Their goal is to create a system that can help patients diagnose diabetes. They used only 4 main algorithms in their research: Decision Tree, SVM algorithm.
The heart plays an important role in humans. Therefore, the prediction of heart disease should be the best and most accurate because heart disease has an important place in humans and can be fatal. So in this article, they describe the accuracy of machine learning algorithms in predicting heart disease. In this study, the authors used k-nearest approximation, decision trees, linear regression and support vector machines .
There are many older studies using different machine learning to predict diabetes, heart disease, kidney disease, Parkinson's disease, accuracy varies depending on the algorithm.

According to this article, diabetes, heart disease, kidney disease, Parkinson's disease, is a serious disease that affects the elderly. In this article, the authors use different machine learning techniques to predict diseases. Their goal is to develop a diabetes, heart disease, kidney disease, Parkinson's disease,predictor that could help the people.

V. RESULTS

Diabetes: SVM

SVM

```
In [70]: from sklearn.svm import SVC
         from sklearn.model_selection import GridSearchCV

         svc = SVC(probability=True)
         parameters = {
             'gamma' : [0.0001, 0.001, 0.01, 0.1],
             'C' : [0.01, 0.05, 0.5, 0.1, 1, 10, 15, 20]
         }

         grid_search = GridSearchCV(svc, parameters)
         grid_search.fit(X_train, y_train)

Out[70]: GridSearchCV(estimator=SVC(probability=True),
                      param_grid={'C': [0.01, 0.05, 0.5, 0.1, 1, 10, 15, 20],
                                  'gamma': [0.0001, 0.001, 0.01, 0.1]})
```

	precision	recall	f1-score	support
0	0.85	0.91	0.88	147
1	0.82	0.72	0.76	81
accuracy			0.84	228
macro avg	0.84	0.81	0.82	228
weighted avg	0.84	0.84	0.84	228

Heart disease: Logistic Regression

Logistic Regression

```
In [26]: accuracies={}

         from sklearn.linear_model import LogisticRegression
         from sklearn.metrics import accuracy_score,confusion_matrix,classification_report
         lr = LogisticRegression(penalty='l2')
         lr.fit(x_train,y_train)

         y_pred = lr.predict(x_test)

         acc=accuracy_score(y_test,y_pred)
         accuracies['LR']=acc*100
         print("Training accuracy score of the model is:",accuracy_score(y_train, lr.predict(x_train))*100,"%")
         print("Testing accuracy score of the model is:",accuracy_score(y_test,y_pred)*100,"%")

Training accuracy score of the model is: 85.49511854951184 %
Testing accuracy score of the model is: 88.31168831168831 %
```

	Disease Name	Algorithm Name	Existing System accuracy	Proposed System accuracy
1	Diabetes	SVM	86%	78%
2	Heartdisease	Logistic Regression	80%	85%
3	Parkinson’sdiseas e	Logistic Regression	72%	87%
4	Kidneydisease	KNN	75%	97%

```
Confusion matrix of the model [[121  24]
 [ 12 151]]
Classification Report              precision    recall  f1-score   support

           0       0.91      0.83      0.87       145
           1       0.86      0.93      0.89       163

    accuracy                           0.88       308
   macro avg       0.89      0.88      0.88       308
weighted avg       0.88      0.88      0.88       308
```

Parkinson’s disease: Logistic Regression

LogisticRegression

```
In [31]: logistic_model = LogisticRegression()

In [32]: logistic_model.fit(x_train, y_train)

Out[32]: LogisticRegression()

In [33]: x_train_prediction = logistic_model.predict(x_train)
         training_data_accuracy = accuracy_score(y_train, x_train_prediction)

In [34]: print("Accuracy on Training Data :", training_data_accuracy)
```

```
Accuracy on Training Data : 0.8717948717948718

Classification Report
              precision    recall  f1-score   support

           0       0.80      0.40      0.53        10
           1       0.82      0.97      0.89        29

    accuracy                           0.82        39
   macro avg       0.81      0.68      0.71        39
weighted avg       0.82      0.82      0.80        39

Confusion Matrix
[[ 4  6]
 [ 1 28]]
```

Kidney disease: KNN

KNN

```
In [42]: from sklearn.neighbors import KNeighborsClassifier
         from sklearn.metrics import accuracy_score, confusion_matrix, classification_report

         knn = KNeighborsClassifier()
         knn.fit(X_train, y_train)

         # accuracy score, confusion matrix and classification report of knn

         knn_acc = accuracy_score(y_test, knn.predict(X_test))

         print(f"Training Accuracy of KNN is {accuracy_score(y_train, knn.predict(X_train))}")
         print(f"Test Accuracy of KNN is {knn_acc} \n")

         print(f"Confusion Matrix :- \n{confusion_matrix(y_test, knn.predict(X_test))}\n")
         print(f"Classification Report :- \n {classification_report(y_test, knn.predict(X_test))}")
```

```
Training Accuracy of KNN is 0.7571428571428571
Test Accuracy of KNN is 0.7

Confusion Matrix :-
[[51 21]
 [15 33]]

Classification Report :-
              precision    recall  f1-score   support

           0       0.77      0.71      0.74        72
           1       0.61      0.69      0.65        48

    accuracy                           0.70       120
   macro avg       0.69      0.70      0.69       120
weighted avg       0.71      0.70      0.70       120
```

VII. CONCLUSION

We wrapped up our conversation by investigating various scholarly papers exhibiting the viability of different models and innovations in diagnosing diabetes, heart disease, kidney disease, Parkinson's disease etc. These thinks about utilized differing machine learning calculations such as back vector machine, Calculated Relapse, K-nearest neighbor, arbitrary woodland, Tensor Flowandkeras .

To guarantee the most recent data, available through eminent databases such as Science Coordinate, IEEE Xplore, Elsevier, and Inquire about Door. In spite of the fact that a few considers depended on their possess information, the larger part utilized datasets from dependable sources counting Cleveland, the UCIcollection, and Lister Slope National Center for Biomedical Communications.

VI. REFERENCES

[1]S. Pouriyeh, S. Vahid, G. Sannino, G. De Pietro, H. Arabnia, and J. Gutierrez, “A comprehensive investigation and comparison of machinelearning techniques in the domain of heart disease,” in 2017 IEEE Symposium on Computers and Communications (ISCC), 2017, pp. 204–207.

[2]Priyanka Sonar, Prof. K. JayaMalini," DIABETES PREDICTION USING DIFFERENT MACHINE LEARNINGAPPROACHES", 2019IEEE, 3rd International Conference on Computing Methodologies and Communication (ICCMC).
[3]Archana Singh, Rakesh Kumar, "Heart Disease Prediction Using Machine Learning Algorithms", 2020 IEEE, International Conference on Electrical and Electronics Engineering (ICE3)
[4]A.Sivasangari, Baddigam Jaya Krishna Reddy,Annamareddy Kiran, P.Ajitha," Diagnosis of LiverDiseaseusingMachineLearningModels"2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)
[5]A. K. Dwivedi, "Performance evaluation of different machine learning techniques for prediction of heart disease," Neural Computing andApplications, vol.29, no. 10, pp. 685– 693, 2018.
[6]K. Polaraju, D. Durga Prasad, and M. Tech Scholar, "Prediction of Heart Disease using Multiple Linear Regression Model," InternationalJournal of Engineering Development and Research, vol. 5, no.4, pp. 2321–9939, 2017. [Online]. Available: www.ijedr.org
[7]Md.NowshadR.Chowdhury,Ezaz Ahmed,Md.Abu Dayan Siddik, Akhlak Uz Zaman, "Heart Disease Prognosis Using Machine Learning Classification Techniques", 2021 6th International Conference for Convergence in Technology (I2CT)https://ieeexplore.ieee.org/document/9418181