IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

MULTIPLE DISEASE PREDICTION SYSTEM USING MACHINE LEARNING

1 Prof. DR. R. Srinivasa Rao, 2 E. Bhargavi, 3 D. Purnachandra Rao, 4 B. Kanthi Kumar, 5 D. Venkatesh

1 Professor, Department of ECE, SRGEC, Gudlavalleru,

2 Undergraduate Student, Department of ECE, SRGEC, Gudlavalleru,

3 Undergraduate Student, Department of ECE, SRGEC, Gudlavalleru,

4 Undergraduate Student, Department of ECE, SRGEC, Gudlavalleru,

5 Undergraduate Student, Department of ECE, SRGEC, Gudlavalleru,

Abstract:

People are becoming increasingly susceptible to high-risk illnesses like chronic diabetes, heart disease, Parkinson’s disease etc. The mortality ratio is prevalent nowadays due to the enormous number of deaths. Many of the current machine learning models for health care analysis focus on a single disease at a time. One analysis is for diabetes, one for the heart, and one for Parkinson's of that nature. There is no standard system that allows one analyst to forecast more than one disease at a time. They also provide a result with low accuracy and precision. Then that lower accuracy puts the patient's life in danger.

Hence, using machine learning we are suggesting a predictive system, the so-called Multiple Disease Prediction System which is used to predict diseases accurately and simultaneously forecast several diseases in one shot. Diabetes, heart disease, and Parkinson's disease are the three diseases we have currently taken into consideration. Future, many more diseases may be added. Here, we employ machine learning (ML) techniques like SVM for (diabetes and Parkinson's) and Logistic Regression for Heart Disease for diseases prediction. The user must enter many disease-related parameters before the system displays a result indicating whether the user has the disease or not.

It aids in the processing of the massive volumes of data produced by the medical sector and significantly shortens the time it takes for doctors to detect a patient's disease at an early stage. It also encourages the adoption of preventative measures to lengthen the patient's life expectancy.

Key Phrases: Machine Learning, Artificial Intelligence, SVM, Logistic Regression, Diabetes, Heart, Parkinson’s.

1. INTRODUCTION

With the advent of the digital era and new technical breakthroughs, necessary data is computerizing for rapid access and helps to transport files to distant distances within seconds. In this digital world, enormous amounts of data are generated day by day and become an asset. So, the medical sector is one of those industries that generates enormous amounts of data. The patient's data about their symptoms, along with clinical factors, hospital resources, diagnostic data, patient’s records, and medical equipment, make up the created data in the health care industry. To extract knowledge for precise decision-making, huge, dense, and complicated data must be processed and assessed. With the help of different algorithms like Decision Tree, Random Forest, SVM, Logistic Regression, and others, medical data mining has a great deal of potential for revealing hidden patterns, correlations, and relationships between features in medical data sets to instantly predict the disease from the patient's symptoms.

Several of the current models focus on just one disease for each analysis. One analysis, for instance, may cover diabetes, one analysis may cover heart, and another analysis for Parkinson's disease. There would be no shared system that could simultaneously analyze multiple diseases.

As a result, we are producing Multiple Disease Prediction System which predicts more than one disease at a time. Here, we are considering giving consumers immediate access to precise disease predictions based on the symptoms they enter. As a result, we are presenting a method that makes use of Streamlit to forecast various diseases. We will examine diabetes, heart disease, and Parkinson's disease analyses in this system. Later, many more illnesses could be added. We will use machine learning algorithms, pickle module, streamlit to implement a multiple disease prediction system. The Python pickling library is used along with the algorithms SVM and Logistic Regression. Model behavior is saved using the Pickling library. An open-source framework called Streamlit is used to create online applications without any prior HTML, CSS, or JavaScript expertise.

The user must supply both the disease name and its parameters when he or she accesses this UI. Invoking the appropriate model, Streamlit will then report the patient's state.

1.1 Description:

Many of the old technologies being used to study data in the medical field only analyze one disease at a time. One system might be used to study diabetes, another to study Parkinson's, and yet another to forecast heart problems. Most of them specialize in a different ailment. A hospital must use more models when it is to test the patient for more disease at a time. A user can concentrate on more diseases using the same webpage using multiple diseases prediction system. To determine whether the user is ill, with the proposed system, the user does not need to travel to several locations. So, the user must choose their favorite disease in MDPS.

1.2 Problem system:

Most of the time, systems utilizing ML can only predict one disease at a time. If we need to test a patient immediately to diagnose him or her for multiple diseases but cannot locate a system that can forecast more than one disease at a time, we will need to use more models, which requires more time and money and has poorer accuracy, endangering the patient's life with false results.

1.3 Proposed System:

Now-a-days Machine learning is rapidly advancing with a variety of algorithms that can process enormous amounts of data, making it possible to automate tasks without the need for human intervention and deliver accurate results right away. Hence, to study many diseases concurrently, we are creating a method called the Multiple Disease Prediction System. The user does not have to visit several websites to research the diseases consequently. For the time being, we have focused on Parkinson's illness, chronic diabetes, and hearing disorders. Later, other illnesses may be added to the MDPS. We are employing ML techniques and Streamlit to implement MDPS. The user must supply the disease parameter along with the disease when accessing the UI.

2. LITERATURE REVIEW

1. The report claims that one of the chronic disorders when blood sugar (glucose) levels are elevated is diabetes. It causes a wide range of illnesses, such as blindness, among others. As it is simple and adaptable to predict if the patient has the condition or not, they have employed ML approaches in the suggested work to analyze diabetes disease. The system's introduction is primarily driven by the need to accurately diagnose diabetes in patients. SVM (Support Vector Machine) and Logistic Regression are the two algorithms employed in the prediction system, and they have accuracy ratings of 78%, 75% respectively. They compared two models' accuracy here [1].

2. The heart plays a crucial role in people, which is the motivation behind the suggested paper. Since heart-related illnesses are on the rise nowadays, it is crucial that they are accurately diagnosed and predicted because they can result in fatal heart complications. Hence, recent developments in AI and ML can enable developing a system that accurately and quickly forecasts the disease. So, using datasets acquired from the well-known website Kaggle, the

authors of this research analyze the accuracy of machine learning (ML) for predicting heart disease using logistic regression, diabetes, and Parkinson's disease using SVM. They also contrasted the methods using SVM (81% and Logistic Regression 82%) accuracy as a benchmark [2].

3. The system targets dopamine deficiency-related Parkinson's disease causes. Dopamine is a vital brain chemical that is thought to help nerves communicate and is known as a neurotransmitter. Parkinson's disease has no known cure, however there are ways to manage the symptoms. Hence, we can quickly identify Parkinson's illness utilizing automated systems and machine learning techniques. For Parkinson's disease prediction, they employed SVM. Accuracy is 87%[3].

3. SYSTEM ANALYSIS

3.1 Functional Requirement:

- The system ought to provide early disease detection by enabling patient disease prediction.
- The user must select the desired ailment from the available list of input values, and the output will be displayed based on the trained model of the user input.

3.2 Non-Functional Requirement:

- A range of values during the prediction and diagnosis will be provided by the website.
- The website needs to be trustworthy and reputable.

4. DESIGN

4.1 Architecture Design:

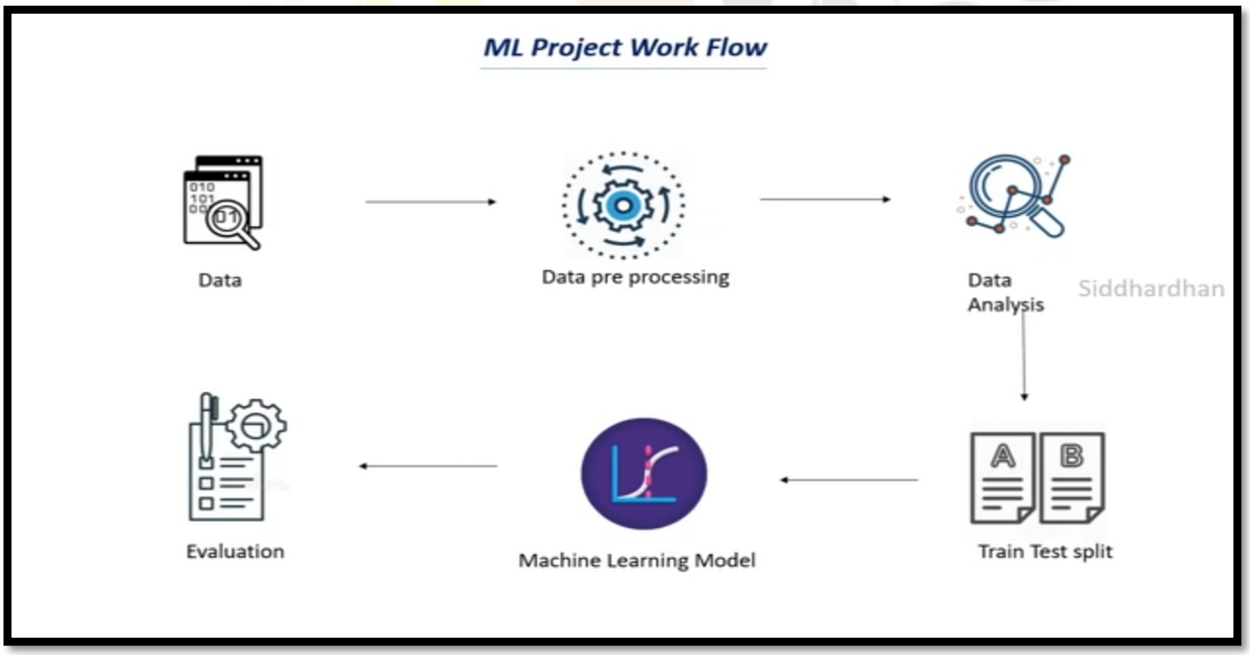

Figure No 4.1: Block Diagram

Fig 4.1., shows architecture diagram for diabetes prediction model. This model has six different modules. These modules include

1. Dataset Collection
2. Data Pre-processing

3. Data Analysis
4. Train Test Split
5. Build Model
6. Evaluation

As diabetes, heart disease, and Parkinson's disease are all related to one another, we conducted experiments on these three disorders in figure 4.1. The dataset for diabetes, heart disease, and Parkinson's disease must be accessed first. The PIMA dataset, the heart disease dataset, and the Parkinson's dataset have all been imported from the Kaggle Website which is the best source for the datasets. After the dataset has been imported, the individual input data is visualized. After data pre-processing for visualization, where we look for outliers, missing values, and scale the dataset, we divide the data into training and testing on the updated dataset.

The training dataset was used to apply SVM and Logistic algorithms, while the testing dataset was used to apply knowledge of the classified algorithm. We will select the algorithm with the highest accuracy for each ailment after applying our knowledge. After that, we create a pickle file for each disease and connect it with the Streamlit framework to provide the model's output on the website.

4.2 User Interface Design:

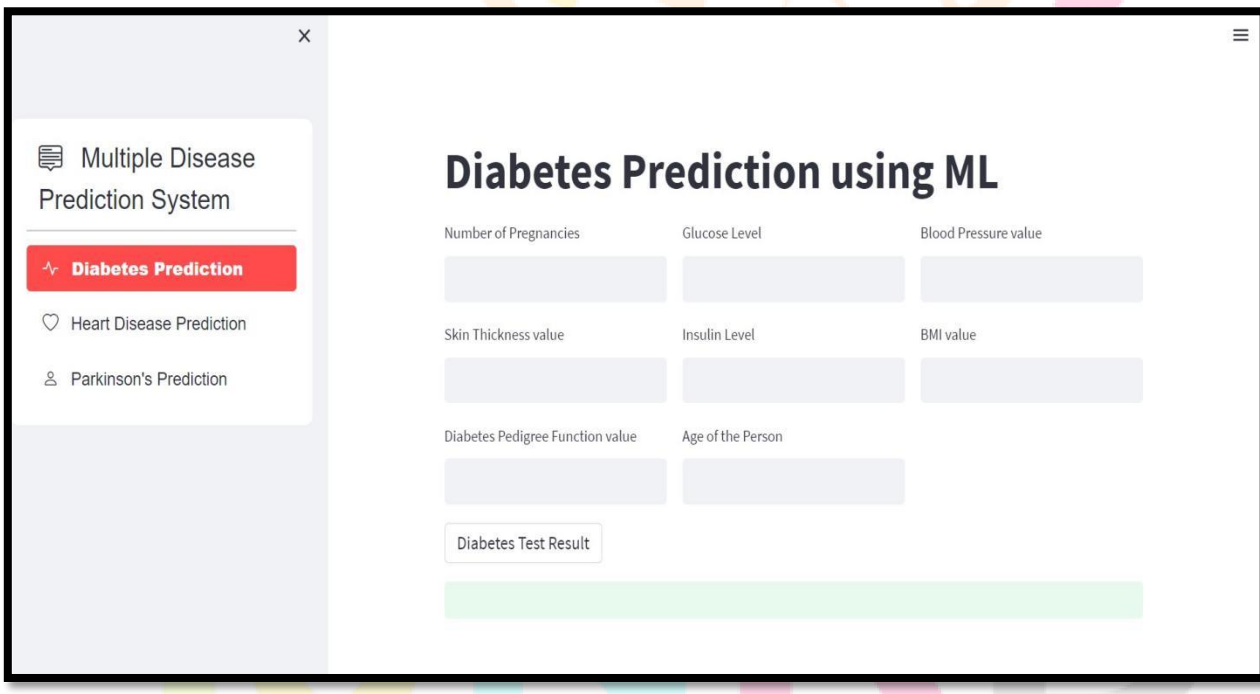

Figure No 4.2: Graphical User Interface

5. IMPLEMENTATION

5.1 Algorithm

5.1.1. Logistic Regression Algorithm:

- Logistic Regression is a supervised learning model
- Classification model
- Uses sigmoid function
- Uses for Binary Classification problems (0 or 1, yes or no, high or low)
- Uses Binary Cross Entropy Loss Function (or) Log Loss Function

Gradient Descent: Gradient Descent is an optimization algorithm used for monitoring the cost function in various machine learning algorithms. It is used for updating the parameters of the learning model

Loss Function: Difference between the estimated value and true value is called Loss Function. Logistic Regression uses Binary Cross Entropy Loss Function (or) Log Loss function.

Binary Cross Entropy Loss Function (or) Log Loss :

$$L(y, \hat{y}) = -(y \log \hat{y} + (1-y) \log(1-\hat{y}))$$

Cost Function: Which deals with a penalty for a number of training sets or the complete batch.

$$J(w, b) = \frac{1}{m}\Sigma(L(y^{(i)}, \hat{y}^{(i)})) = -\frac{1}{m}\Sigma(y^{(i)} \log \hat{y}^{(i)} + (1-y^{(i)}) \log(1-\hat{y}^{(i)}))$$

Operation of Logistic Regression:

Step 1: Adjust the Logistic Regression Classifier to the preprocessed dataset.

Step 2: The classifier fits the dataset to create the prediction model. The best weight and bias parameters are provided by the built model for precise prediction.

Step 3: The following procedures can be used to determine the ideal parameters

Equations for building models:

1. Sigmoid Function:

$$\hat{y} = \frac{1}{1+e^{-Z}} \qquad Z = w.X + b$$

Sigmoid Function

2. Updating weights through Gradient descent:

w2 = w1 – L*dw

b2 = b1 – L*db

3. Derivatives:

dw = 1/m(y_hat-y).X

db = 1/m(y_hat-y)

Step 4: The model has been trained and is now capable of disease prediction.

Step 5: By entering the user parameters and the anticipated disease, the patient's condition can be determined.

w: weight, b: bias, x: input feature, y hat: expected/predicted value (probability of y being 1), L: Learning Rate, Z: Straight line equation, m: Total no of training examples, dw: Partial Derivative of cost function with respect to weight, db: Partial Derivative of cost function with respect to Bias, y: True value, w2: Denotes a new weight, b2: A new bias value, w1: A previous weight and b1: Bias.

Note: The appropriate weight and bias values can be found using the Gradient Descent Algorithm in order to determine the minimum cost function.

5.1.2. Support Vector Machine Algorithm:

- Model for Supervised Learning
- Classification and regression both
- Sufficient for High Dimensional Dataset
- Uses Hinge Loss Function

Hyperplane: A hyperplane is a line or plane that divides the data points into two classes in two-dimensional (2D) space.

Support Vectors: The datapoints that are closest to the hyperplane are support vectors. The hyperplane's position changes if these data points change.

Margin: The distance between two lines on distinct classes' nearest data points.

Maximum margin: An ideal hyperplane is one that has the largest margin.

Equation of Maximum Margin:

$$\max \left(\frac{2}{||w||} \right) \text{ Such that,}$$

$$y_i = \begin{cases} -1, & w^T x_1 + b \leq -1 \\ 1, & w^T x_1 + b \geq 1 \end{cases}$$

Loss Function : Loss function measures how far an estimated value is from its true value. It is helpful to determine which performs better & which parameters are better.

Hinge Loss: Support Vector uses the Hinge Loss function which is one of the types of Loss function, mainly used for maximum margin classification models. Hinge Loss incorporates a margin or distance from the classification boundary into the loss calculation. Even if new observation are classified correctly, they can incur a penalty if the margin from the decision boundary is not large enough

Equation of Hinge Loss Function: L = MAX(0,1 - yi(w.T*xi + b))

Operation of SVM:

Step 1: Fit the Support Vector Machine Classifier to the preprocessed dataset.

Step 2: The classifier fits the dataset to create the prediction model. The best weight and bias parameters are provided by the built model for precise prediction.

Step 3: The following procedures can be used to determine the ideal parameters:

Equations for building models:

1. Equation of the Hyperplane:

y = w*x - b

2. Updating weights through Gradient descent:

w2 = w1 – L*dj/dw

b2 = b1 – L*dj/db

3. Derivatives:

If(yi. (w*x + b) >= 1):

dj/dw = 2*lambda*w

dj/db = 0

else (yi.(w.x+b)<1):

dj/dw = 2*lambda*w – yi.xi

dj/db = yi

Step 4: The model has been trained and is now capable of disease prediction.

Step 5: By entering the user parameters and the anticipated disease, the patient's condition can be determined.

w: weight, b: bias, x: input feature, y hat: expected/predicted value, L: Learning Rate, dj/dw: Partial Derivative of cost function with respect to weight, dj/db: Partial Derivative of cost function with respect to Bias, y: true value, lambda: regularization parameter, w2: denotes a new weight, b2: a new bias value, and w1: a previous weight and b1: old bias.

Note: The appropriate weight and bias values can be found using the Gradient Descent Algorithm in order to determine the minimum cost function.

6. RESULT

Three diseases are included in the Multiple Disease Prediction System: diabetes, heart disease, and Parkinson's. SVM is used to predict diabetes and Parkinson's disease, and logistic regression is used to predict heart disease for greater accuracy. The system returns the patient's state once the user provides the relevant parameters and the projected disease. The warning indication that adds a correct value will appear if the value fields are out of range or empty.

ACCURACY FOR EACH DISEASE:

Table No 6.1: Diabetes Disease

ALGORITHM	DIABETES_DISEASE
Support Vector Machine	78%
Logistic Regression	75%

Table No 6.2: Heart disease

ALGORITHM	HEART_DISEASE
Support Vector Machine	81%
Logistic Regression	82%

Table No 6.3: Parkinson's disease

ALGORITHM	PARKINSON'S_DISEASE
Support Vector Machine	87%
Logistic Regression	87%

1. Diabetes Disease:

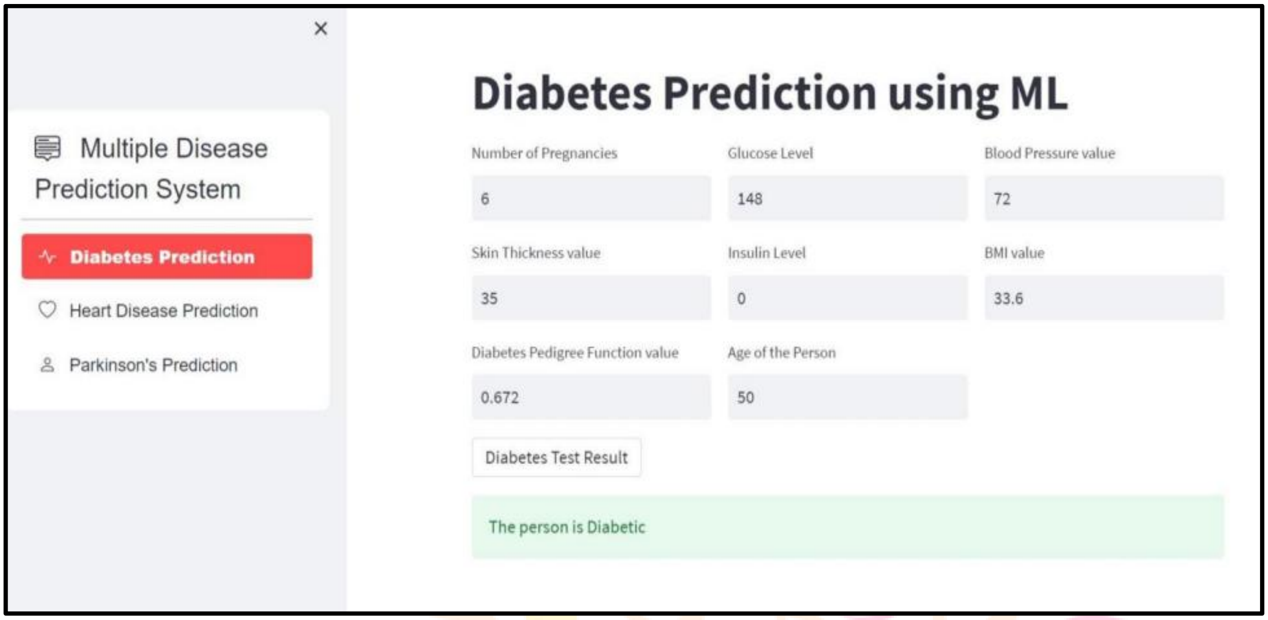

Figure No 6.1: The person is Diabetic

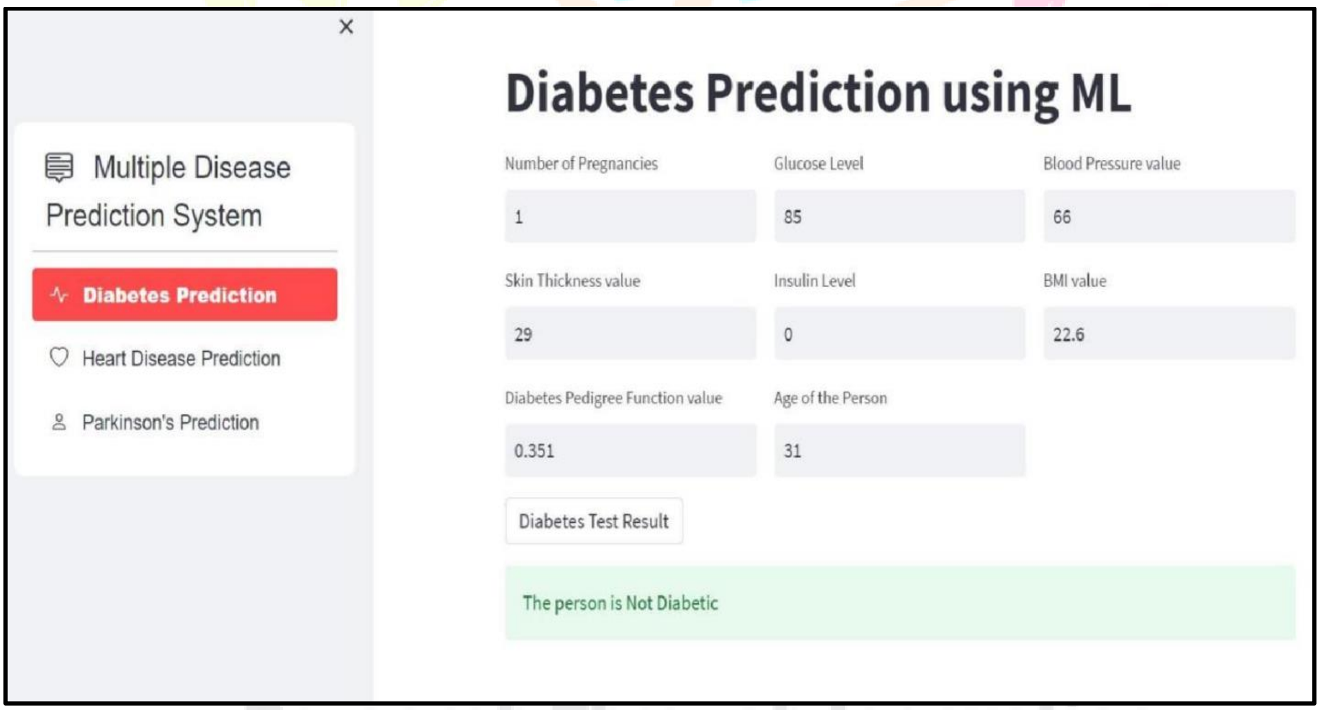

Figure No 6.2: The person is Not Diabetic

2. Heart Disease:

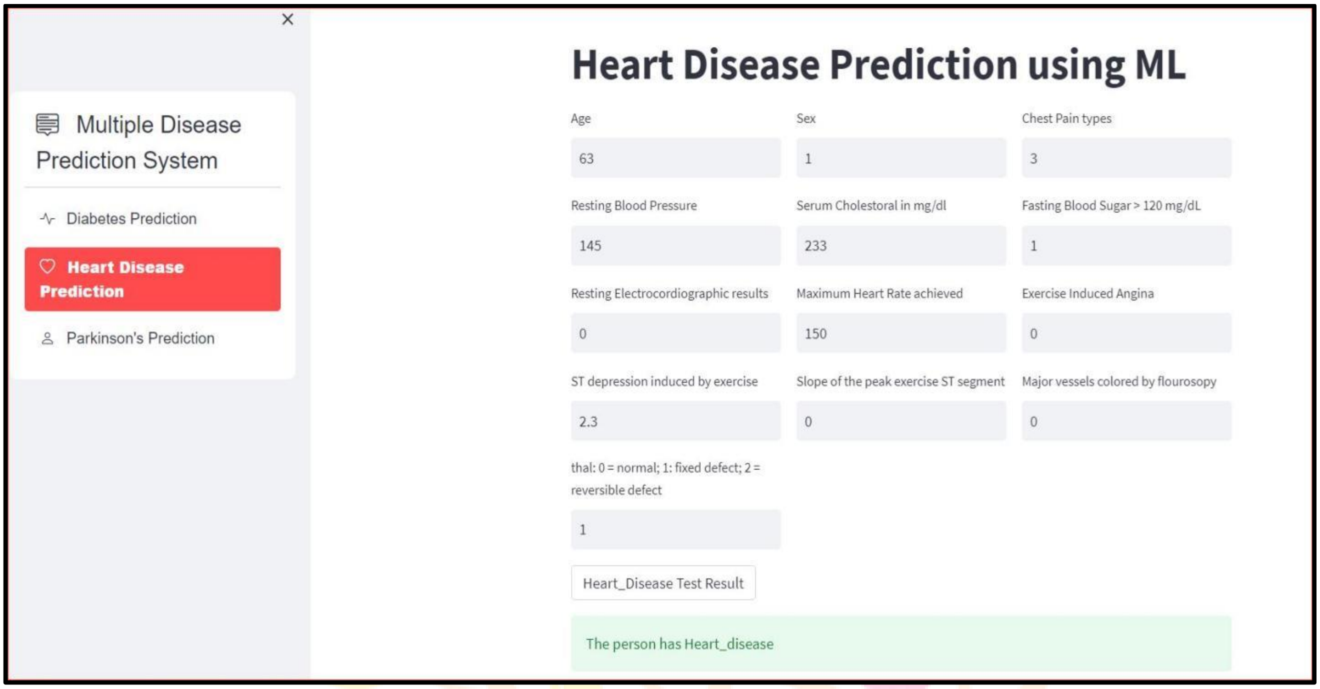

Figure No 6.3: The person has heart disease

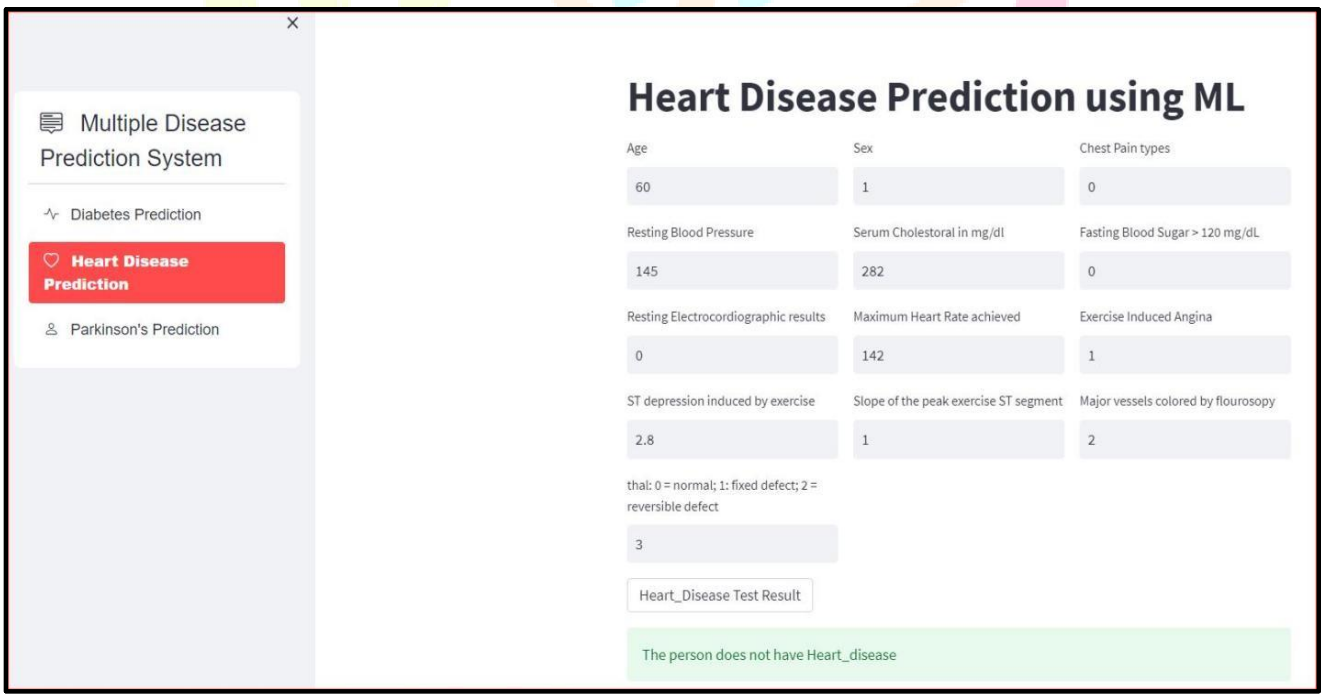

Figure No 6.4: The person does not have heart disease

3. Parkinson's Disease:

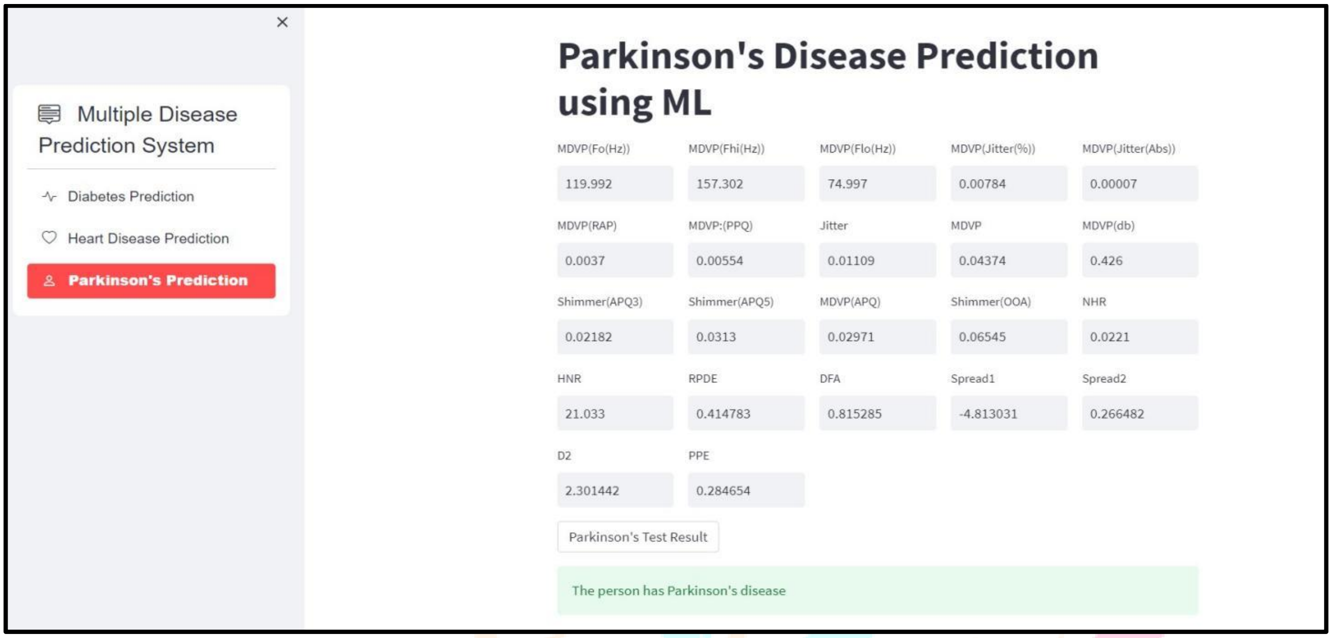

Figure No 6.5: The person has Parkinson's disease

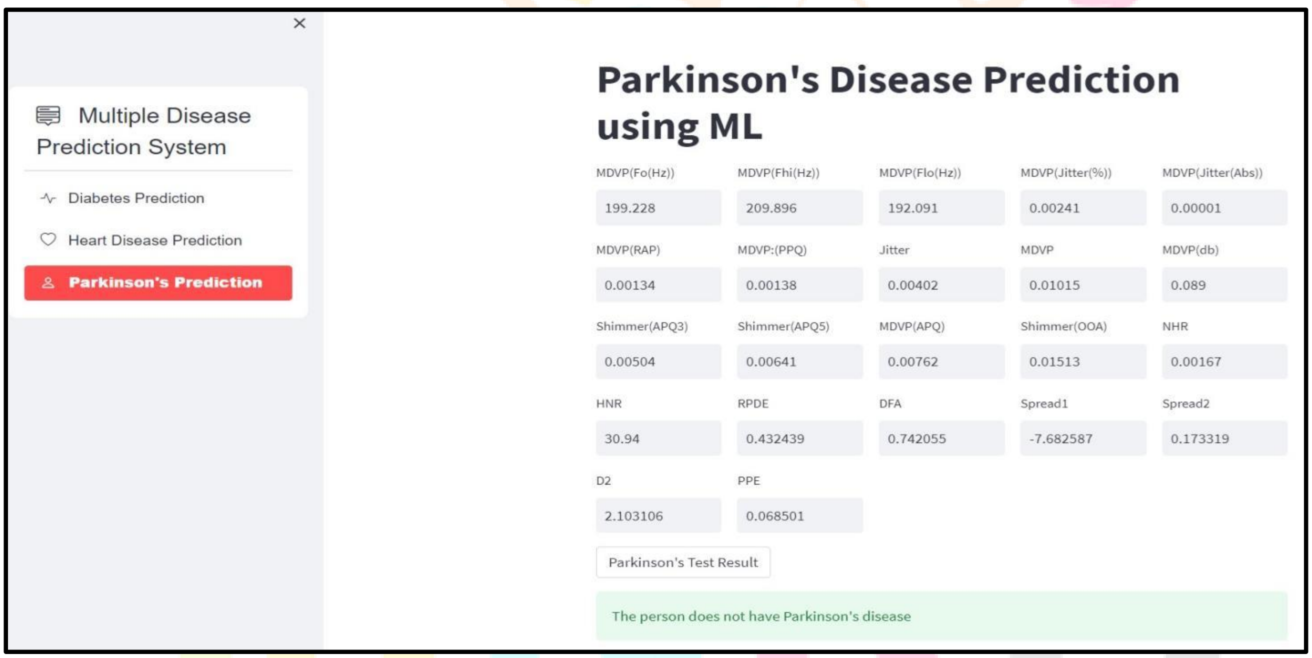

Figure No 6.6: The person does not have Parkinson's disease

7. CONCLUSION

The goal of this research was to develop a multi-disease prediction system that would provide accurate predictions right away. This project allows users to handle enormous amounts of data more quickly and without having to navigate through numerous websites, both of which add to the speed of prediction. In addition to increasing human life expectancy and preventing financial hardship, early disease prediction can help us take preventative actions, which in turn lowers the mortality ratio. To achieve the highest level of accuracy when compared to other classification algorithms like decision tree, Naive Bayes, Random Forest, KNN, we used machine learning algorithms like Support Vector Machine and Logistic Regression for implementing Multiple Disease Prediction System.

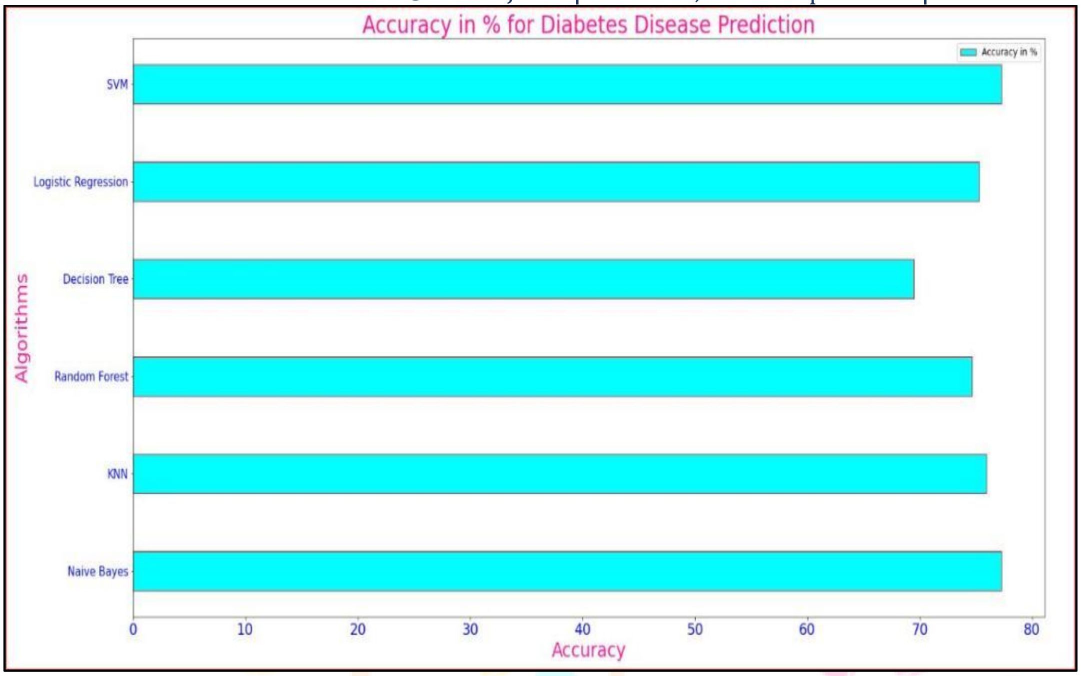

Figure No 7.1: Results of classification algorithms for predicting diabetes disease accuracy

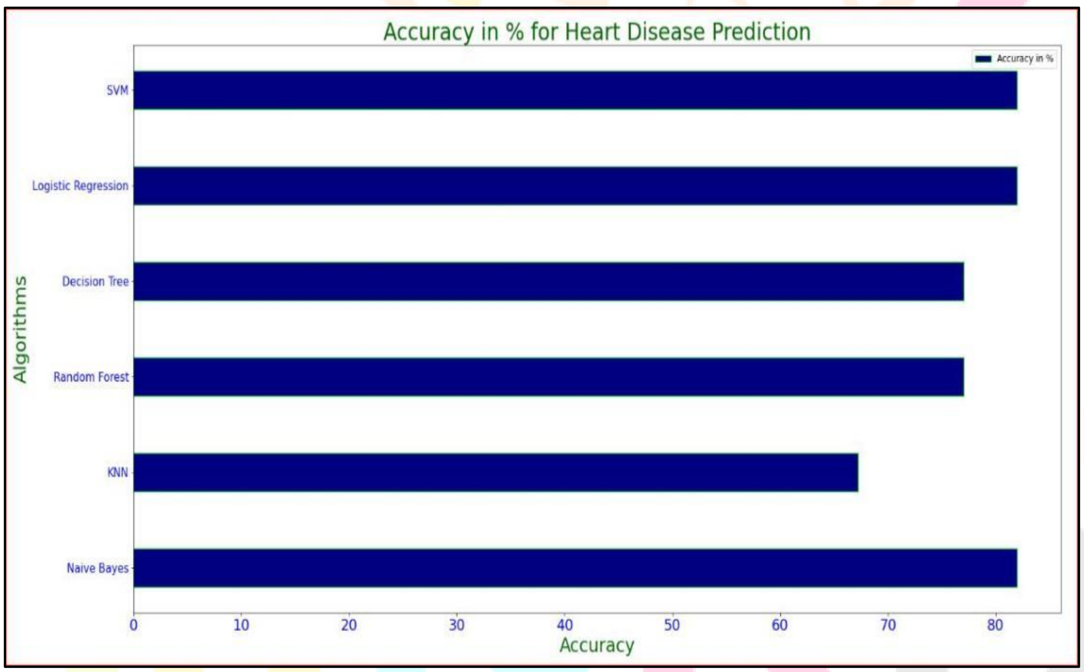

Figure No 7.2: Results of classification algorithms for predicting heart disease accuracy

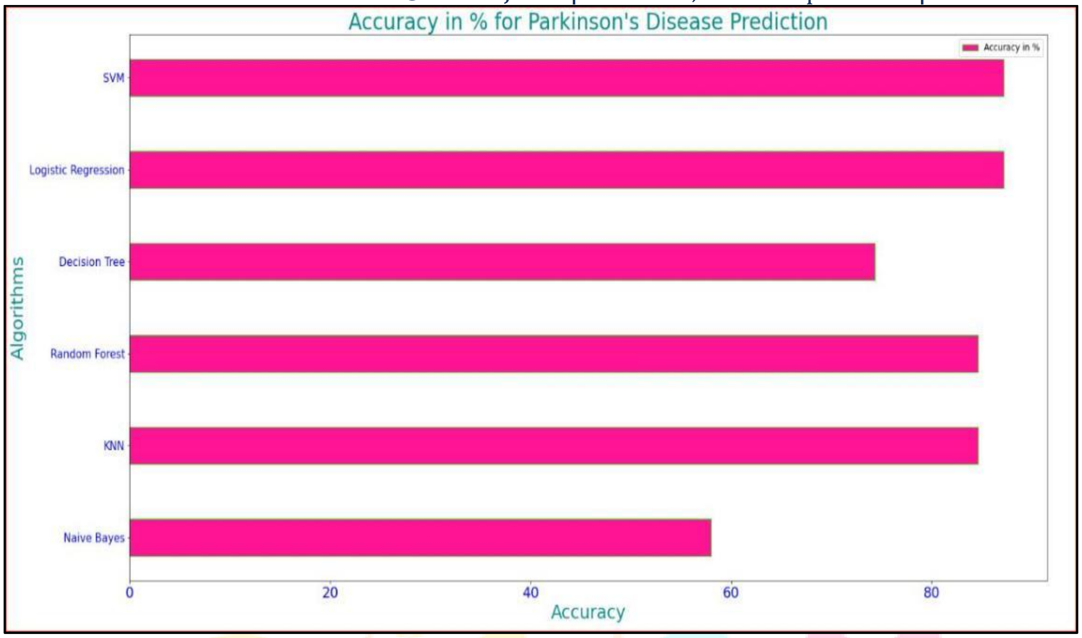

Figure No 7.3: Results of classification algorithms for predicting parkinson's disease accuracy

8. FUTURE SCOPE

- The current User Interface could eventually be expanded to include a considerable number of ailments.
- To extend life expectancy, we can work to boost prediction accuracy.
- Try to provide a dependable, approachable, and consistent user interface that includes illness stages and the essential safety measures based on that disease degree.

ACKNOWLEDGEMENT

We would like to express our gratitude to Dr. G. V. S. N. R. V. Prasad, the principal of Seshadri Rao Gudlavalleru Engineering College, for providing us with the opportunity and time to conduct and study on the topic of "Multiple Disease Prediction System Using Machine Learning." We appreciate DR. R. Srinivasa Rao, who served as our mentor, and Prof. Y. Rama Krishna, the department's head of electronics and communication, for their assistance during our research, which would have seemed challenging without their inspiration, ongoing support, and insightful suggestions. Without the cooperation, advice, and assistance of our friends and family, the complexity of this study article would not have been achievable.

REFERENCES

[1] Laxmi Deepthi Gopisetti, Srinivas Karthik Lambavai Kummera, Sai Rohan Pattamsetti, Sneha Kuna, Niharika Parsi, Hari Priya Kodali, "Multiple Disease Prediction Model by using Machine Learning and Streamlit" 2023 IEEE, 5th International Conference on Smart Systems and Inventive Technology (ICSSIT)

[2] Akkem Yaganteeswarudu, "Multi Disease Prediction Model by using Machine Learning" 2020 IEEE, 5th International Conference on Communication and Electronics Systems (ICCES)

[3] Elsevier B.V," Diabetes Prediction Using Machine Learning" 2019, International Conference on Recent Trends in Advanced Computing.

[4] KM Jyoti Rani, "Diabetes Prediction Using Machine Learning" July 2020, International Journal of Scientific Research in Computer Science Engineering, and Information Technology

[5] Firdous, Shimoo, Wagai, Gowher A, Sharma, Kalpana, “A survey on diabetes risk prediction using machine learning approaches”, November 2022, *Journal of Family Medicine, and Primary Car.*

[6] Krittanawong, C. Virk, H. U., Bengaluru, S., Wang, Z., Johnson, K. W., Pinotti, R., Zhang, H., Kaplin, S., Narasimhan, B., Kitai, T., Baber, U., Halperin, J. L., & Tang, W. H. (2020). Machine learning prediction in cardiovascular diseases.

[7] Chaimaa Boukhatem, Heba Yahia Youssef, Ali Bou Nassif. February 2022 IEEE, Advances in Science and Engineering Technology International Conferences (ASET)

[8] Supriya Kamoji, Dipali Koshti, Valiant Vincent Dmello, Alrich Agnel Kudel, Nash Rajesh Vaz, Prediction of Parkinson's Disease using Machine Learning and Deep Transfer Learning from different Feature Sets, July 2021 IEEE, 6th International Conference on Communication and Electronics Systems (ICCES).

[9] Rohit Surya, A.T., Yaswanthram, P., Nair, P.R., Rajendra Prasath, S.S., Akella, S.V. (2022). Prediction of Parkinson’s Disease Using Machine Learning Models—A Classifier Analysis. In: Bianchini, M., Piuri, V., Das, S., Shaw, R.N. (eds) Advanced Computing and Intelligent Technologies. Lecture Notes in Networks and Systems, vol 218. Springer, Singapore. https://doi.org/10.1007/978-981-16-2164-2_35.

[10] Makarious, M. B., Leonard, H. L., Vitale, D., Iwaki, H., Sargent, L., Dadu, A., Violich, I., Hutchins, E., Saffo, D., Kim, J. J., Song, Y., Maleknia, M., Bookman, M., Nojopranoto, W., Campbell, R. H., Hashemi, S. H., Botia, J. A., Carter, J. F., Craig, D. W., . . . Nalls, M. A. (2022). Multi-modality machine learning predicting Parkinson’s disease. *Npj Parkinson's Disease*, *8*(1), 1-13. https://doi.org/10.1038/s41531-022-00288-w.