VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

School of Computer Science and Engineering

Programme : B.Tech

Course Title : Machine Learning

Course Code : BCSE209L

Slot : G2

Title: Parkinson's and Alzheimer's disease detection using Machine learning algorithms

Team Members: Khushi Agarwal | 21BCE1459

Rupkatha De | 21BCE6075

Yash Bawa | 21BCE1670

Faculty: Dr. Rajalakshmi R **Sign:**

Table of Contents

Abstract

Parkinson's disease (PD) is a widespread neurological condition that has a significant global impact on individuals. According to reports, PD affects an estimated 10 million people worldwide (0.3% of the total population), with those over 60 making up 1% of those affected.

The most common neurodegenerative condition is Alzheimer's disease. Although the signs are first mild, they worsen over time. Among dementias, Alzheimer's disease is one of the most common. There is no cure for this disease, making it difficult to treat. The sickness is identified, but only at an advanced stage. Therefore, if the condition is identified earlier, its course or symptoms may be slowed.

We are proposing a system which can predict the Parkinson's and Alzheimer's disease based on the test records obtained from the clinic

The importance of this project that many people across the globe suffer with either of these diseases and getting to know about these diseases in the earlier stages might help them recover with some treatment. So, building a model which can accurately predict the condition of the patient is very crucial.

The results of this project will be telling us if the person has Parkinson's disease or not and for the Alzheimer's case it will tell what kind of class that like Non-Demented, Very Mild Demented, Mild Demented, Moderately Demented.

When it comes to the comparison of results with the existing method, we provided a better accuracy model for predicting the Parkinson's disease using different machine learning classifiers. We also used CNN for classifying the images for Alzheimer's which works better than SVM and Decision tree classifiers.

Keywords

Linear Regression

Support Vector Machine

Convolutional Neural network

Random Forest

Decision Tree

Naïve Bayes

K nearest Neighbours

Introduction

The substantia nigra pars compacta in the midbrain is where dopaminergic neurons in Parkinson's disease (PD) die. A variety of symptoms, such as stiffness, bradykinesia, voice alterations, and coordination problems are brought on by this neurodegeneration. Patients with Parkinson's disease (PD) may also experience dysarthria, which is characterised by motor-speech system weakness, paralysis, and loss of coordination that affects respiration, phonation, articulation, and prosody. Since the disease's course and symptoms can vary, PD is frequently misdiagnosed for many years. New sensitive diagnostic methods are therefore required for PD identification since, as the disease worsens, more symptoms appear that make PD more challenging to treat.

Loss of intensity, monotony of pitch and loudness, diminished stress, inappropriate silences, brief bursts of speech, variable tempo, incorrect consonant articulation, and a harsh and breathy voice are the main deficiencies of PD speech (dysphonia). Due to the non-invasive nature of capturing voice data and the ease with which it may be done using mobile devices, the variety of voice-related symptoms appears promise for a prospective screening tool.

Alzheimer's disease (AD) is a form of brain malfunction where the patient's mental abilities gradually deteriorate. Memory loss and dementia are its most evident symptoms. About 35 million individuals were affected by AD in 2010, and it's predicted that 1 in 85 people would get the disease by 2050. Usually, memory loss develops gradually. Short-term memory and learning are the only memory functions first affected, but with time, long-term memory would also suffer. According to researchers, AD begins years before any symptoms appear.

Researchers are looking for methods to identify AD in persons at risk early because it takes a long time before clinical symptoms appear. In order to prevent or at least slow the progression of this disease, it is necessary to identify anatomical alterations in the brains of those who are at high risk for it. The size of the brain's regions can alter and be detected by magnetic resonance imaging (MRI). It is possible to use the measurement of the ROIs where atrophy develops during AD progression as a diagnostic sign.

Many studies have been conducted on the diagnosis of AD using brain MRIs, however only a small portion of the entire brain has been observed in these studies, and the remaining portions have not yet been researched. The majority of research have used volume image analysis-based ROIs in multiple MRI slices to identify the condition. This will make it more difficult for the research's findings to be used in MRI facilities. The objective of this study was to determine whether it was possible to distinguish between AD and HC participants with a level of accuracy that was acceptable utilising just one slice based on pixel colour.

Literature survey

[1] Parkinson's Disease Diagnosis Using Machine Learning and Voice

- In this research paper, they could diagnose and predict the Parkinson's disease through machine learning architectures using only non-invasive voice biomarkers as features. Because of its underlying cognitive and neuromuscular function, biomarkers extracted from human voice can provide insight into neurological illnesses like Parkinson's disease (PD).
- Advantage of this research paper is that they used different machine learning algorithms such as Decision tree, SVM, Random Forest etc on the given dataset so that even if one of these models is not very good with the following dataset, the other models might work well and predict the outcomes with a better accuracy.
- Disadvantages in this paper is that even though they used different machine learning models, they couldn't reach an accuracy score of 90 percent. The highest they gained was 86 percent and also the data used for these algorithm's performance was limited to a clinician. Few studies suggest that voice analysis is not that accurate to predict the Parkinson's disease.
- The overcoming solution we have come up with to increase the accuracy of the model is, we have used many more other classifiers and models like KNN, Linear regression model, Naïve Bayes, which weren't used in the following results. Some of these showed a better accuracy than the rest for our dataset and also, we split our dataset into 90 training and 10 testing which gave the model more info to be trained and predict the outcomes more accurately.

[2] A Novel Method for Parkinson's Disease Diagnosis Utilizing Treatment Protocols

- In this research paper they collected the data from the University of Baghdad's College of Medicine's treatments Protocol Repository. They did survey on both female and male participants. In this they asked the patients to repeat the vowel 'a' thrice and collected their voice modulations and the frequency and took their previous medical conditions and also took their blood reports.
- Advantage of this research paper is unlike other research papers the data they used is more diverse. They collected the data from patients of different genders, different ages groups and few control subjects who doesn't have Parkinson disease. They also used different machine learning algorithms to check whichever algorithm gives the better outputs.
- Disadvantages in this paper is their data set is very small. They conducted a survey among only 253 participants and the they divided the dataset for both testing and training. The accuracy obtained from these models are low. The average accuracy they were receiving was 65 percentage.
- The overcoming solution we have come up with to increase the accuracy of the model is, we have used the dataset which is as diverse as this dataset but also contains more data samples compared to this which helped our model to give a better accuracy.

[3] Diagnosis of Parkinson's disease using EEG and fMRI

- In this research paper, they collected the fMRI and EEG data from the patients and used that to diagnose and predict the Parkinson's disease. Here the collected data is in the form of images. fMRI scans are of the scans of their brain and the EEG produces the readings of the electrical activity that is taking place in the brain.
- Advantages of this paper is that, they used the scans of brain and like many studies suggested that brain scans give more insights about the person's Parkinson condition than the data collected from that same person's vocal modulation and frequencies.
- Disadvantages of this paper is that, collecting brain scans survey is difficult because the scans are costlier than the voice recordings. Due to this reason the data available for this paper was less compared to the other papers.
- The overcoming solution we have come up for this is that we can just go with the voice frequency since we have enough data and even if we need some extra data, we can collect it from people around us just by taking their voice frequencies.

[4] Prediction of Alzheimer's disease using Machine learning Classifiers.

- In this research paper, they took the MRI scans of brain and used those images as their dataset to train the models which will help in detecting the Alzheimer's disease.
- Advantage of this paper is that for detecting Alzheimer's the only way is to use MRI scans of the brain and for machine learning algorithms they used classifiers like Support Vector Machine and Bayesian Support Vector Machines. Based on the accuracy that each model is giving, they proceeded with the one giving the most accuracy.
- Disadvantage of this paper is that they used SVM. SVM gives good results but for SVM they first had to create some features from the image and then they have to feed those features to the classifier and also there might be human errors while creating those features and there will be chances of losing spatial interaction between pixels during this process.
- The overcoming solution we have come up for this is that we used CNN model instead of SVM. CNN can be taught of automatic feature extraction which will reduce the possible errors that can occur when we make our own features and feed the model.

[5] Alzheimer Disease Prediction using Machine Learning Algorithms

- In this research paper, they took the MRI scans of brain and used those images as their dataset to train the models which will help in detecting the Alzheimer's disease.
- Advantage of this paper is that for detecting Alzheimer's the only way is to use MRI scans of the brain and for machine learning algorithms they used different classifiers like Support Vector Machine and Decision tree classifier. Based on the accuracy that each model is giving, they proceeded with the one giving the most accuracy.
- Disadvantage of this paper is that they used SVM and decision tree. Both SVM and Decision tree works very good and give better results when it comes to the numerical parameters but for these models they first had to create some features

from the image and then they have to feed those features to the classifier and also there might be human errors while creating those features and there will be chances of losing spatial interaction between pixels during this process.

- The overcoming solution we have come up for this is that we used CNN model instead of SVM or decision tree classifier. CNN can be taught of automatic feature extraction which will reduce the possible errors that can occur when we make our own features and feed the model.

[6] Deep learning in Alzheimer's disease: Diagnostic classification and prognostic prediction using Neuroimaging data

- They have conducted an eligibility screening for over 192 publications and concluded from the filtered 16 papers that they will be making use of multimodal neuroimaging data to identify structural and molecular/functional biomarkers for Alzheimer's.
- To select optimal features from multimodal neuroimaging data for diagnostic classification, there is usually need of several pre-processing steps such as neuroimaging registration and feature extraction, which greatly affects the classification performance. However, in this model, deep learning approaches have been applied to AD diagnostic classification using original neuroimaging data without any feature selection procedures.
- Since they have went with one particular approach to learn the data, there is no show of accuracy between different models being applied to the data.
- We have made use of all the prominent learning algorithms in our model. Further, based on this, we have used the model with the higher accuracy.

[7] Alzheimer's detection based on segmentation of an MRI image

- They have used the system of Magnetic Resonance Imaging (MRI) images and applying the model of wavelet transform to detect Alzheimer's.
- Their proposed method involves applying wavelet transform on the input image and then using inverse wavelet transform. Next, they proceed to reconstruct the image using 3D reconstruction and then segment it. Based on this the decision is made.
- Since wavelet transform has been used here, some of the serious disadvantages of this model are: shift sensitivity, poor directionality and lack of phase information.
- Our model is based on choosing the best possible model by testing it with different learning methods and choosing the one with the best accuracy and hence there are no cases of overfitting or shift sensitivity.

[8] Alzheimer detection using Group Grey Wolf Optimization based features with convolution classifier

- This paper employed the text information for features' extraction in addition to the segmentation of brain images.
- The model they have implemented enhanced the binary and multi0 class classification execution. The multi-class classification process remains the commitment of this work which is enabled with the chosen ideal features from Group Grey Wolf Optimization (GGWO) - motivated enhancement model.

- Some of the critical flaws of this model are that it has a lack of accuracy, low population diversity, premature convergence and imbalance between the exploitation and exploration.
- Our employed model overcomes flaws such as population diversity as our input data is voice recordings of patients and thus applying the model with the highest accuracy on this would yield the best possible results.

[9] Detecting Parkinson's disease with sustained phonation and speech signals using machine learning techniques

- The main contribution of this approach is evaluating the performance with the use of other classifiers and processing audio without fusing of feature sets.
- Their proposed approach is based on motor symptoms explicitly related to voice. The voice examination begins with two vocal tasks namely - phonation and speech respectively. Based on the data they selected the best configuration for each classifier.
- This model employed by them is expensive in terms of time since they have used varied parameters for their dataset which will not be a time efficient solution.
- Our model provides a more time efficient solution since it focuses on the important parameters in the voice recordings that are necessary for the learning models in order to detect the disease.

[10] A deep learning approach for Parkinson's disease diagnosis from EEG signals

- Electroencephalogram (EEG) signals have been used as input data in this model since this is a disease related to brain abnormality. PD is characterized by gradual degradation of the motor function in the brain.
- In this work, they have used EEG signals of twenty PD and twenty normal subjects. This model has achieved a promising performance of 88.25% accuracy, 84.71% sensitivity, and 91.77% specificity.
- Since this model involves processing EEG signals as input data it is extremely expensive in terms of processing speeds and power. This model is allso not ready to be installed for clinical usage.
- Our input data is a collection of thousands of MRI scans of the brain which is used by the test data later on to compare with. We have employed an automated detection system for PD using the convolutional neural network (CNN).

[11] Parkinson's progression prediction using machine learning and serum cytokines.

- This paper talks about how the model uses Serum Cytokines samples from patients with Parkinson's Disorder with and without the common leucine-rich repeat kinase 2 (LRRK2) G2019S mutation and determines the progression rate.
- An advantage of this model is that it produces high accuracy results because it uses the serum samples.
- But one drawback is collecting these samples is very difficult and sometimes expensive to collect.
- In our project, we simply make use of the voice frequency values to determine the Parkinson's disorder and it is more easily accessible.

[12] Early-Stage Alzheimer's Disease Prediction Using Machine Learning Models.

- This paper discusses about how it identifies Alzheimer's disease in early stages. Predictions of Alzheimer's disease are based on Open Access Series of Imaging Studies (OASIS) data, and performance is measured with parameters like Precision, Recall, Accuracy, and F1-score for ML models. These datasets are available on Kaggle. The proposed classification scheme can be used by clinicians to make diagnoses of these diseases.
- An advantage of this is that it uses high accuracy Machine Learning models and is very efficient.
- Our project uses MRI images dataset to determine the Alzheimer's disease.

[13] Parkinson's disease prognostic scores for progression of cognitive decline.

- This model uses analysis of 19 baseline clinical, pathological and demographic variables.
- An advantage of this model is that they employ a varied set of parameters for this data.
- A drawback, since it has so many parameters, it is computationally expensive and leads to overfitting.
- Our model is more efficient because it uses MRI images which are classified and easy to train.

[14] A Comparative Analysis of Machine Learning Algorithms to Predict Alzheimer's Disease

- This paper uses the OASIS dataset to determine the Alzheimer's disease.
- An advantage of this is that it uses different attributes and gets more accurate results.
- Our project uses MRI images to determine Alzheimer's disease

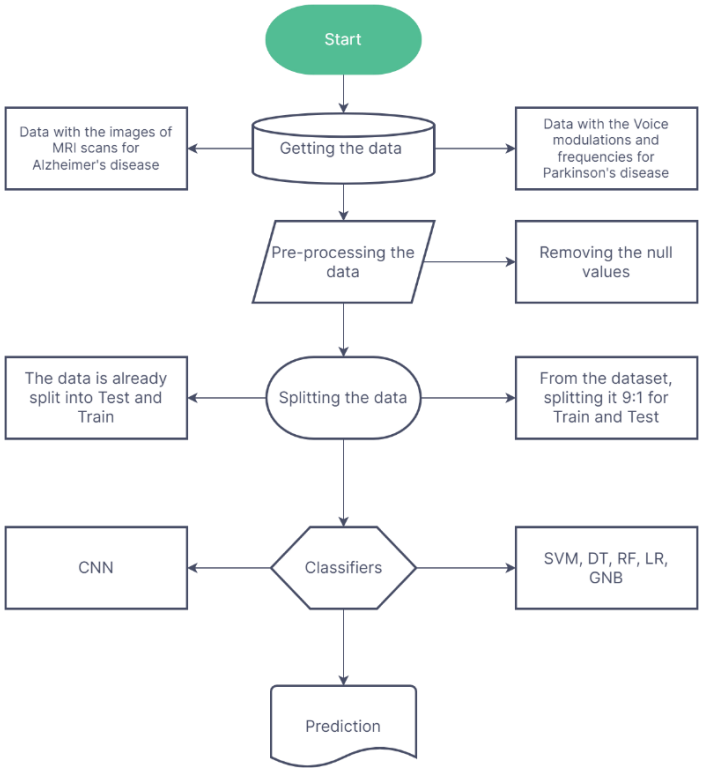
[15] Optimizing Machine Learning Methods to Improve Predictive Models of Alzheimer's Disease.

This paper discusses about investigating the accuracy of different ML methods and different features to classify cognitively normal (CN) individuals from Alzheimer's disease (AD) and to predict longitudinal outcome in participants with mild cognitive impairment (MCI).

Proposed Method

In our project we tried to predict both Alzheimer's and Parkinson's disease using various Machine Learning Classifiers. For Parkinson's detection, we got the dataset from Kaggle which has the data of voice modulation and frequencies of different patients belonging to both the genders and age from 33 to 87 years. We used Linear Regression, Support Vector Machine, Decision Tree, Random Forest, Naïve Bayes, K nearest neighbours' classifiers to check which algorithm gives us the most accuracy for this particular dataset after splitting into both testing and training dataset. We observed that the KNN gave us the best accuracy of 98 percent when the value of K is 3 but its accuracy fell to 96 percent when the value of K got changed to 5. The next best model was Decision tree which gave us a solid 97 percent accuracy. So, we moved forward with the Decision tree Classifier and predicted the outcomes for the given test dataset.

For Alzheimer's disease prediction we again got the data from the Kaggle. Alzheimer's data was more organised than the Parkinson's data. The dataset was divided into Train and Test and each had 4 sub divisions named Non-Demented, Mild Demented, Very Mild Demented, Moderately Demented. There were a combined of 6400 images of Brain MRI scans. These MRI scans are used to predict the Alzheimer's disease. We used the Convolutional Neural Network to predict. The reason for using this model over SVM and Decision tree is, in SVM and DT, we have to create our own feature vectors and have to feed the model. During the feeding process there is a possibility of loss in the spatial interaction between the pixel. Whereas in CNN the machine can be taught in automated feature extraction from the image.



Experimental Results and Discussions

Importance of Parameters taken into consideration and the software testing done on dataset.

For Parkinson's disease prediction we took a dataset that consists the records from both male and female patients and also the age group varies from 33 to 87 years. The attributes that are taken into the consideration are

MDVP:Fo(Hz) - Average vocal fundamental frequency
MDVP:Fhi(Hz) - Maximum vocal fundamental frequency
MDVP:Flo(Hz) - Minimum vocal fundamental frequency
MDVP:Jitter(%),MDVP:Jitter(Abs),MDVP:RAP,MDVP:PPQ,Jitter:DDP - Several measures of variation in fundamental frequency
MDVP:Shimmer,MDVP:Shimmer(dB),Shimmer:APQ3,Shimmer:APQ5,MDVP:APQ,Shimmer:DDA - Several measures of variation in amplitude
NHR,HNR - Two measures of ratio of noise to tonal components in the voice
status - Health status of the subject (one) - Parkinson's, (zero) - healthy
RPDE,D2 - Two nonlinear dynamical complexity measures
DFA - Signal fractal scaling exponent
spread1,spread2,PPE - Three nonlinear measures of fundamental frequency variation.

The raw data collection contains 753 features that can be linked to certain feature groups. Wavelet features and MFCCs (mel-frequency cepstral coefficients), in addition to basic qualities like intensity parameters and formant frequencies (formal phonetic frequencies), are some of the features that can be discovered (wavelet properties). The three categories were determined to be combined due to the similarities in the traits of the new time features (intensity, formant frequency, and bandwidth parameters).

For Alzheimer's disease prediction we took the image data set which consists of MRI Brain scans of different patients that are suffering with Alzheimer's and also took the scans of the controlled patients who are not diagnosed with Alzheimer's disease. The software testing on the dataset to give the important parameters is given below

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 176, 176, 16)	448
conv2d_1 (Conv2D)	(None, 176, 176, 16)	2320
max_pooling2d (MaxPooling2D)	(None, 88, 88, 16)	0
sequential (Sequential)	(None, 44, 44, 32)	14016
sequential_1 (Sequential)	(None, 22, 22, 64)	55680
sequential_2 (Sequential)	(None, 11, 11, 128)	221952
dropout (Dropout)	(None, 11, 11, 128)	0
sequential_3 (Sequential)	(None, 5, 5, 256)	886272
dropout_1 (Dropout)	(None, 5, 5, 256)	0
flatten (Flatten)	(None, 6400)	0
sequential_4 (Sequential)	(None, 512)	3279360
sequential_5 (Sequential)	(None, 128)	66176
sequential_6 (Sequential)	(None, 64)	8512
dense_3 (Dense)	(None, 4)	260

Total params: 4,534,996
Trainable params: 4,532,628
Non-trainable params: 2,368

Results obtained by our work in terms of accuracy, runtime, precision, recall.

For Parkinson's

Logistic Regression

********** LogisticRegression(C=0.4, max_iter=1000, solver='liblinear') ********

	precision	recall	f1-score	support
0	0.72	0.75	0.73	24
1	0.82	0.80	0.81	35
accuracy			0.78	59
macro avg	0.77	0.78	0.77	59
weighted avg	0.78	0.78	0.78	59

Confusion Matrix:
[[18 6]
 [7 28]]

Decision Tree

```
********** DecisionTreeClassifier(random_state=14) *********
              precision    recall  f1-score   support

           0       1.00      0.92      0.96        24
           1       0.95      1.00      0.97        35

    accuracy                           0.97        59
   macro avg       0.97      0.96      0.96        59
weighted avg       0.97      0.97      0.97        59

Confusion Matrix:
[[22  2]
 [ 0 35]]
```

Random Forest

```
********** RandomForestClassifier(random_state=14) *********
              precision    recall  f1-score   support

           0       1.00      0.92      0.96        24
           1       0.95      1.00      0.97        35

    accuracy                           0.97        59
   macro avg       0.97      0.96      0.96        59
weighted avg       0.97      0.97      0.97        59

Confusion Matrix:
[[22  2]
 [ 0 35]]
```

Random Forest (Entropy)

```
********** RandomForestClassifier(criterion='entropy') *********
              precision    recall  f1-score   support

           0       1.00      0.92      0.96        24
           1       0.95      1.00      0.97        35

    accuracy                           0.97        59
   macro avg       0.97      0.96      0.96        59
weighted avg       0.97      0.97      0.97        59

Confusion Matrix:
[[22  2]
 [ 0 35]]
```

Support Vector Machine

```
********** SVC(cache_size=100) *********
              precision    recall  f1-score   support

           0       0.92      0.92      0.92        24
           1       0.94      0.94      0.94        35

    accuracy                           0.93        59
   macro avg       0.93      0.93      0.93        59
weighted avg       0.93      0.93      0.93        59

Confusion Matrix:
[[22  2]
 [ 2 33]]
```

K nearest neighbours

```
********** KNeighborsClassifier(n_neighbors=3) *********
              precision    recall  f1-score   support

           0       0.96      0.96      0.96        24
           1       0.97      0.97      0.97        35

    accuracy                           0.97        59
   macro avg       0.96      0.96      0.96        59
weighted avg       0.97      0.97      0.97        59

Confusion Matrix:
[[23  1]
 [ 1 34]]
```

Naïve Bayes

```
********** GaussianNB() *********
              precision    recall  f1-score   support

           0       0.88      0.88      0.88        24
           1       0.91      0.91      0.91        35

    accuracy                           0.90        59
   macro avg       0.89      0.89      0.89        59
weighted avg       0.90      0.90      0.90        59

Confusion Matrix:
[[21  3]
 [ 3 32]]
```

For Alzheimer's

Convolution Neural Network

	precision	recall	f1-score	support
NonDemented	0.97	1.00	0.98	639
VeryMildDemented	1.00	1.00	1.00	635
MildDemented	0.93	0.92	0.92	662
ModerateDemented	0.92	0.91	0.91	624
micro avg	0.95	0.95	0.95	2560
macro avg	0.95	0.95	0.95	2560
weighted avg	0.95	0.95	0.95	2560
samples avg	0.95	0.95	0.95	2560

Result Tabulation

Dataset Features:

```
     MDVP:Fo(Hz)  MDVP:Fhi(Hz)  MDVP:Flo(Hz)  MDVP:Jitter(%)  \
0        119.992       157.302        74.997         0.00784
1        122.400       148.650       113.819         0.00968
2        116.682       131.111       111.555         0.01050
3        116.676       137.871       111.366         0.00997
4        116.014       141.781       110.655         0.01284
..           ...           ...           ...             ...
190      174.188       230.978        94.261         0.00459
191      209.516       253.017        89.488         0.00564
192      174.688       240.005        74.287         0.01360
193      198.764       396.961        74.904         0.00740
194      214.289       260.277        77.973         0.00567

     MDVP:Jitter(Abs)  MDVP:RAP  MDVP:PPQ  Jitter:DDP  MDVP:Shimmer  \
0             0.00007   0.00370   0.00554     0.01109       0.04374
1             0.00008   0.00465   0.00696     0.01394       0.06134
2             0.00009   0.00544   0.00781     0.01633       0.05233
3             0.00009   0.00502   0.00698     0.01505       0.05492
4             0.00011   0.00655   0.00908     0.01966       0.06425
..                ...       ...       ...         ...           ...
190           0.00003   0.00263   0.00259     0.00790       0.04087
191           0.00003   0.00331   0.00292     0.00994       0.02751
192           0.00008   0.00624   0.00564     0.01873       0.02308
193           0.00004   0.00370   0.00390     0.01109       0.02296
194           0.00003   0.00295   0.00317     0.00885       0.01884

     MDVP:Shimmer(dB)  ...  MDVP:APQ  Shimmer:DDA      NHR     HNR      RPDE  \
0               0.426  ...   0.02971      0.06545  0.02211  21.033  0.414783
1               0.626  ...   0.04368      0.09403  0.01929  19.085  0.458359
2               0.482  ...   0.03590      0.08270  0.01309  20.651  0.429895
3               0.517  ...   0.03772      0.08771  0.01353  20.644  0.434969
4               0.584  ...   0.04465      0.10470  0.01767  19.649  0.417356
..                ...  ...       ...          ...      ...     ...       ...
190             0.405  ...   0.02745      0.07008  0.02764  19.517  0.448439
191             0.263  ...   0.01879      0.04812  0.01810  19.147  0.431674
192             0.256  ...   0.01667      0.03804  0.10715  17.883  0.407567
193             0.241  ...   0.01588      0.03794  0.07223  19.020  0.451221
194             0.190  ...   0.01373      0.03078  0.04398  21.209  0.462803

          DFA   spread1   spread2        D2       PPE
0    0.815285 -4.813031  0.266482  2.301442  0.284654
1    0.819521 -4.075192  0.335590  2.486855  0.368674
2    0.825288 -4.443179  0.311173  2.342259  0.332634
3    0.819235 -4.117501  0.334147  2.405554  0.368975
4    0.823484 -3.747787  0.234513  2.332180  0.410335
..        ...       ...       ...       ...       ...
190  0.657899 -6.538586  0.121952  2.657476  0.133050
191  0.683244 -6.195325  0.129303  2.784312  0.168895
192  0.655683 -6.787197  0.158453  2.679772  0.131728
193  0.643956 -6.744577  0.207454  2.138608  0.123306
194  0.664357 -5.724056  0.190667  2.555477  0.148569

[195 rows x 22 columns]
```

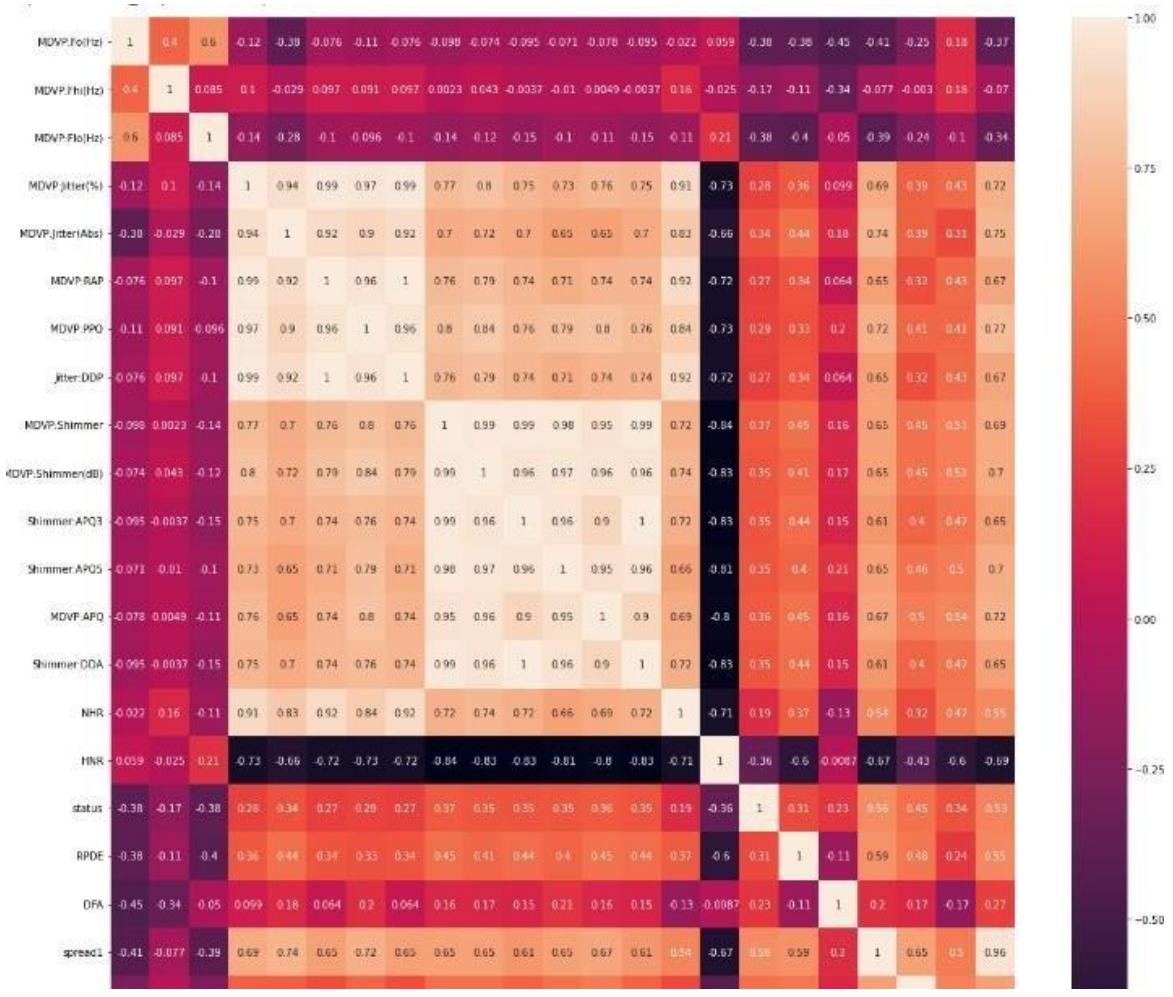
Dataset Labels:

```
0      1
1      1
2      1
3      1
4      1
      ..
190    0
191    0
192    0
193    0
194    0
Name: status, Length: 195, dtype: int64
```

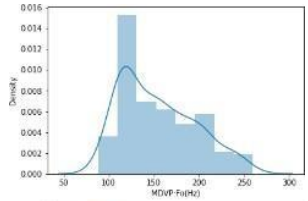
Comparative analysis in form of graph

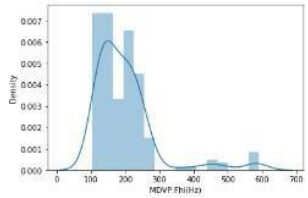
Parkinson

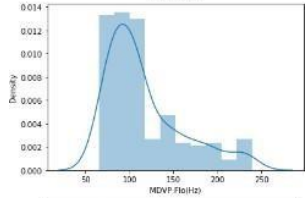
Corelation

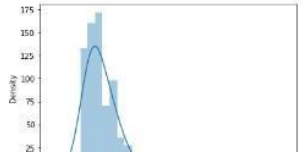


Density

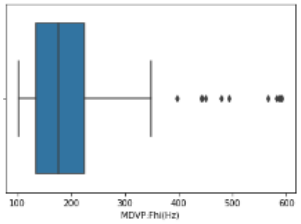


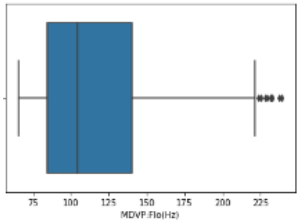


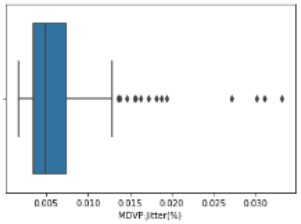


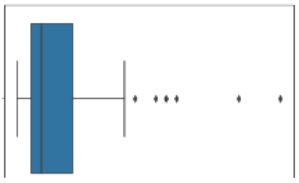


BoxPlot

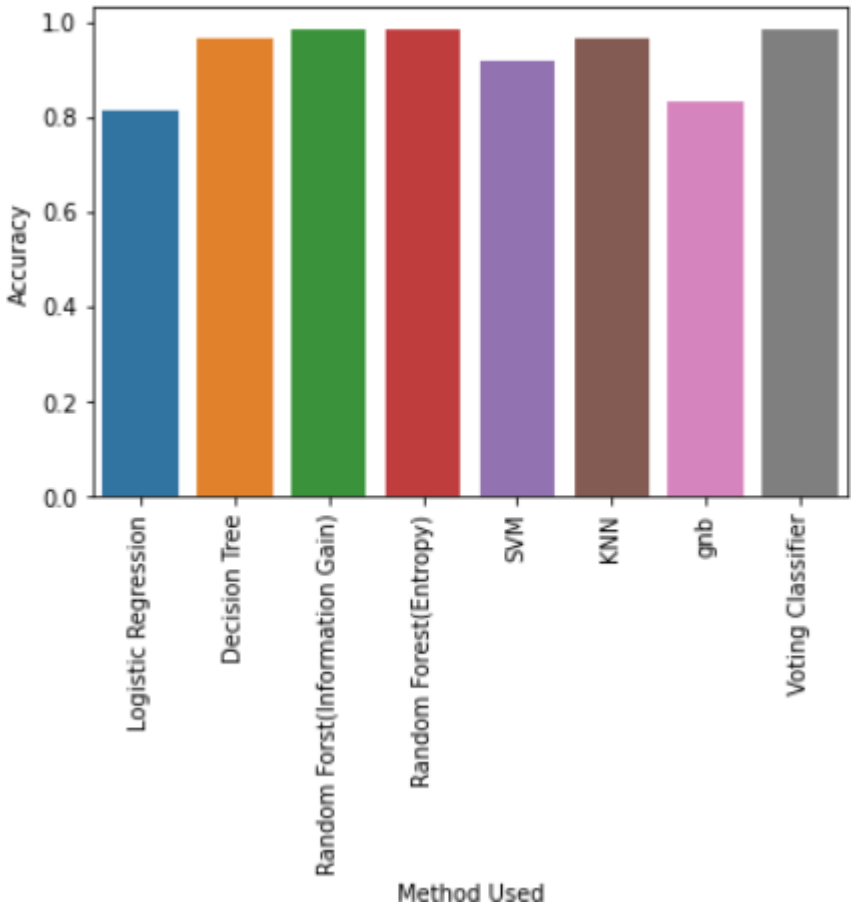








Our model



Model from the literature survey

(a)

Data set / Performance criteria	Baseline data group			MFCC data group		
	Classifier algorithms					
	Decision tree (DT)	kNN	SVM	Decision tree (DT)	kNN	SVM
Accuracy rate (%)	71.35	68.75	66.67	69.79	75	73.44
Sensitivity	0.75	0.7	0.72	0.6	0.72	0.7
Specificity	0.68	0.68	0.61	0.79	0.79	00.78
F-measurement	0.72	0.69	0.66	0.69	0.75	0.73
Kappa	0.43	0.38	0.33	0.4	0.5	0.47
Area under the ROC curve	0.71	0.71	0.68	072	0.72	0.72

(b)

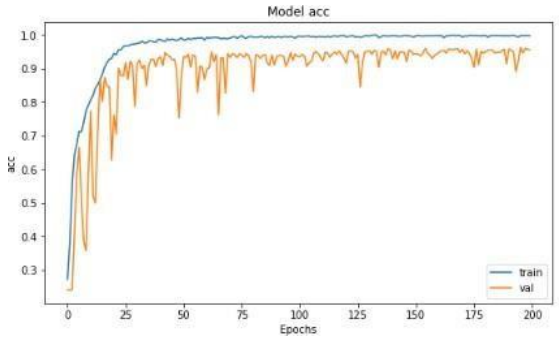
Data set / Performance criteria	Time data group			Vocal data group		
	Classifier algorithms					
	Decision tree (DT)	kNN	SVM	Decision tree (DT)	kNN	SVM
Accuracy rate (%)	66.15	67.71	72.4	61.98	64.06	63.02
Sensitivity	0.73	0.82	0.8	0.64	0.58	0.64
Specificity	0.59	0.53	0.65	0.6	0.7	0.63
F-measurement	0.65	0.65	0.72	0.62	0.64	0.63
Kappa	0.32	0.35	0.45	0.24	0.28	0.26
Area under the ROC curve	0.65	0.69	0.70	0.61	0.63	0.64

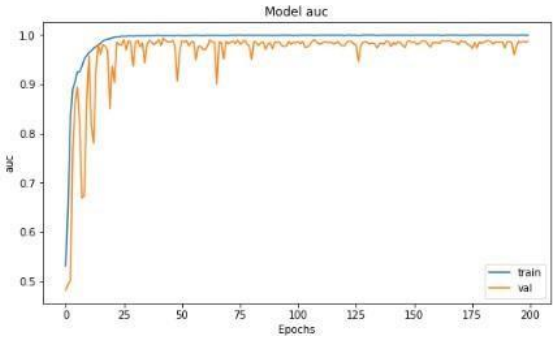
(c)

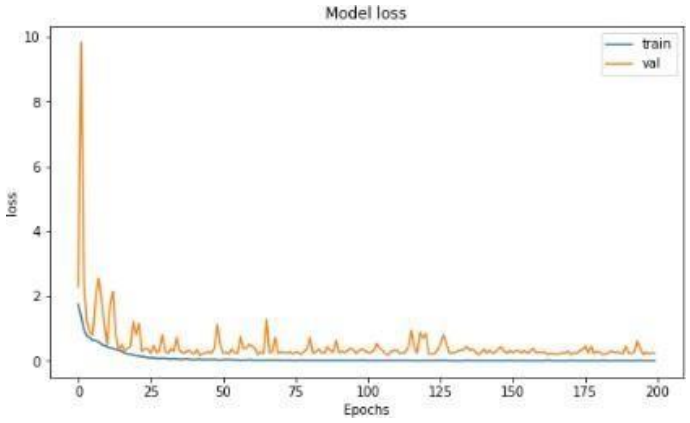
Data set / Performance criteria	Wavelet data group			Whole data group		
	Classifier algorithms					
	Decision tree (DT)	kNN	SVM	Decision tree (DT)	kNN	SVM
Accuracy rate (%)	68.23	73.44	79.69	69.79	76.56	85.42
Sensitivity	0.81	0.76	0.97	0.6	0.81	0.94
Specificity	0.55	0.72	0.63	0.79	0.72	00.78
F-measurement	0.66	0.73	0.76	0.69	0.76	0.86
Kappa	0.36	0.47	0.59	0.4	0.53	0.72
Area under the ROC curve	0.69	0.74	0.82	0.70	0.78	0.89

Alzheimer's

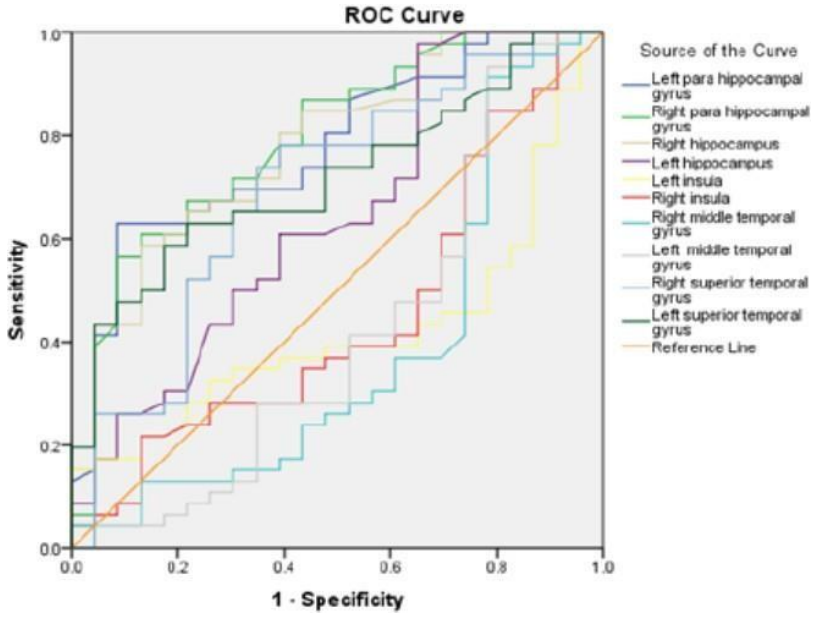
Our model







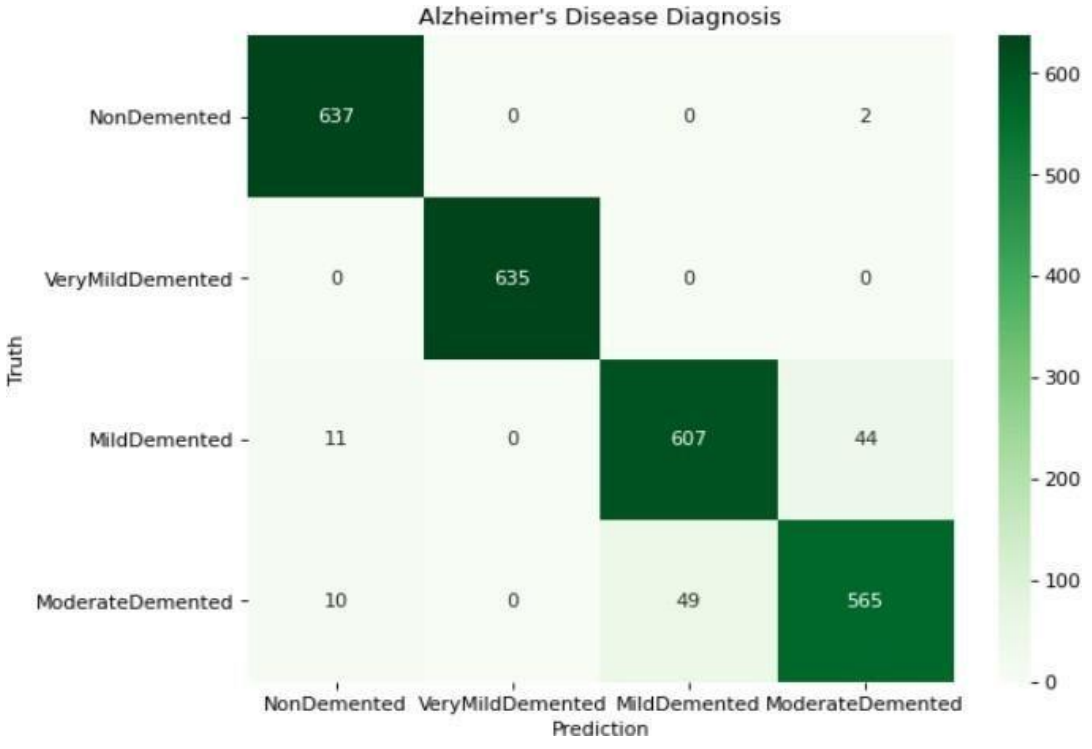
Model from Literature survey



Conclusion

For predicting the Parkinson's disease using only non-invasive vocal biomarkers as features, automated machine learning systems were successfully implemented to diagnose and forecast disease. With noisy and high dimensional data, our investigation compares how well different machine learning classifiers diagnose diseases. Clinical level precision is feasible with careful feature selection. These results are encouraging because they could pave the way for brand-new ways to use voice data to evaluate patient health and neurological conditions. There is reason to expect that in the future, denser feature sets containing spoken word, video, or other modalities would help in illness prediction and clinical confirmation of diagnosis because of the great accuracy demonstrated by the models using these brief audio samples.

The one more goal of this project was to use brain MRI to distinguish between Alzheimer's and healthy control participants using Convolution Neural Networks. The project's findings can be effectively put into practise. The results' degree of accuracy using the Convolutional neural network classifier reached 95 percent.



References

https://www.researchgate.net/publication/328726100_Parkinson%27s_Disease_Diagnosis_Using_Machine_Learning_and_Voice

https://www.researchgate.net/publication/362433153_A_Novel_Method_for_Parkinson's_Disease_Diagnosis_Utilizing_Treatment_Protocols

https://www.researchgate.net/publication/364338000_Diagnosis_of_Parkinson's_disease_using_EEG_and_fMRI

https://www.researchgate.net/publication/344766443_Prediction_of_Alzheimer%27s_Disease_Using_Machine_Learning_Classifiers

https://ieeexplore.ieee.org/document/9074248

Deep learning in Alzheimer's disease: Diagnostic classification and prognostic prediction using Neuroimaging data (Published: 2019) Taeho Jo, Kwangsik Nho, Andrew J. Saykin

Alzheimer's detection based on segmetation of an MRI image (Published: 2017) K.S.Biju, S.S.Alfa, Kavya Lal, Alvia Antony, M Kurup Akhil

Alzheimer detection using Group Grey Wolf Optimization based features with convolution classifier (Published: 2019) Shankar K, Lakshmanaprabu S.K., Ashis Khanna, Sudeep Tanwar, Joel J.P.C. Rodrigues, Nihar Raqnjan Roy

Detecting Parkinson's disease with sustained phonation and speech signals using machine learning techniques (Published: 2019) Jefferson S. Almeida, Pedro P. Reboucas Filho, Tiego Carneiro, Wei Wei, Robertas Damsevicius, Rytis Maskeliuna, Victor Hugo C.de Albuquerque

A deep learning approach for Parkinson's disease diagnosis from EEG signals (Published: 2018) Shu Li Oh, Yuki Hagiwara, U.Raghavendra, Rajamanickam Yuvaraj, N. Arunkumar, M.Murugappan, U. Rajendra Acharya

https://doi.org/10.1038/s41531-019-0086-4

https://doi.org/10.3389/fpubh.2022.853294

https://doi.org/10.1038/s41598-019-54029w

https://doi.org/10.1155/2021/9917919

https://doi.org/10.3233/JAD-190262