Introduction to Machine Learning (CSL2010)
Group Project Report
Topic: Handwriting for Alzheimer's Detection

Siddhartha Singh (B22BB041)

Ishaan Pandey (B22CI017)

G Dileep (B22CH011)

Yash Kumar (B22MT047)

Deekshant Singh Rajawat (B22CI010)

Novemeber 2023

Project code link

Contents

1 Introduction

1.1 Background

- Alzheimer's disease is responsible for gradual decline in cognitive function, memory loss, and changes in behaviour.
- People suffering from Alzheimer's have profound impacts on the lives and their families. Early diagnosis is crucial for effective intervention.
- Subtle changes in handwriting patterns, such as alterations in size, speed, and pressure, can serve as potential indicators of cognitive decline.
- In this project, we are provided with a dataset containing 450 features extracted from handwriting's of unaffected and affected people.

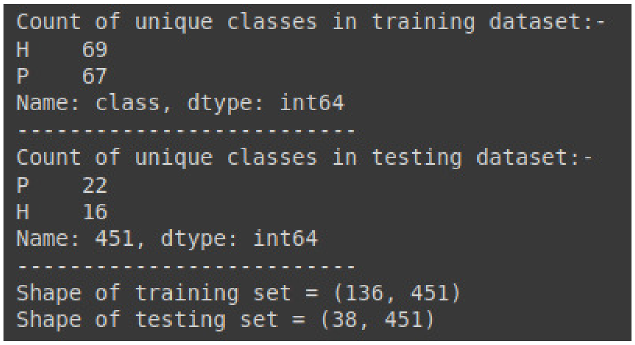
1.2 Objective

- This project aims to explore the application of **machine learning**, specifically analysis of handwriting features, for the detection of Alzheimer's disease.

- We seek to create a machine learning model with appreciable accuracy that can aid in the early diagnosis of Alzheimer's, allowing for timely intervention and personalized care.

2 Preprocessing & EDA

- We have used Standard Scaler for normalization of input features.
- We have checked the distribution of classes in the training and testing set. Snapshot is attached below :-

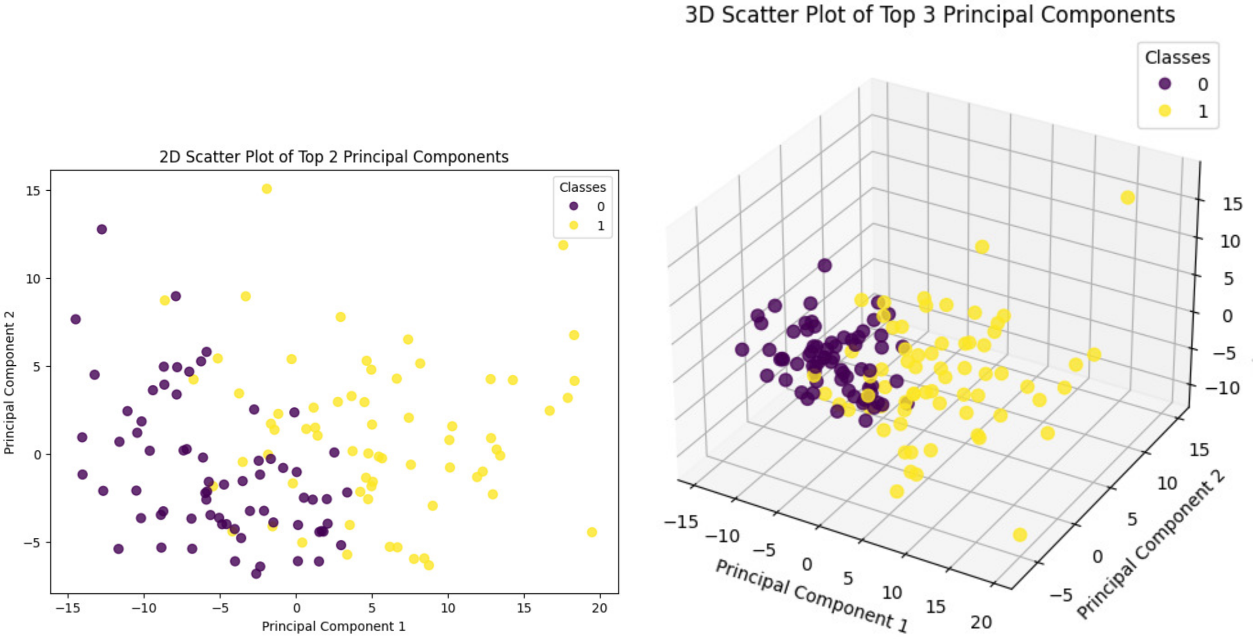


```
Count of unique classes in training dataset:-
H    69
P    67
Name: class, dtype: int64
--------------------------
Count of unique classes in testing dataset:-
P    22
H    16
Name: 451, dtype: int64
--------------------------
Shape of training set = (136, 451)
Shape of testing set = (38, 451)
```

Figure 1: Count of classes labelled as P in training set = 67 and in testing set = 22
Count of classes labelled as H in training set = 69 and in testing set = 16

3 Principle Component Analysis

- We did feature extraction using Principal Component Analysis (PCA). As PCA assumes that data is captured in the variance of features.
- We captured 80% variance and reduced the number of features from more than 400 to just 48.
- We visualized the top 2 and top 3 principal components using matplotlib. The plots are as shown :-



((a)) Top 2 PC's

((b)) Top 3 PC's

- We also plotted the cumulative and individual explained variance graphs for our 48 principal components. The graph is shown below :-

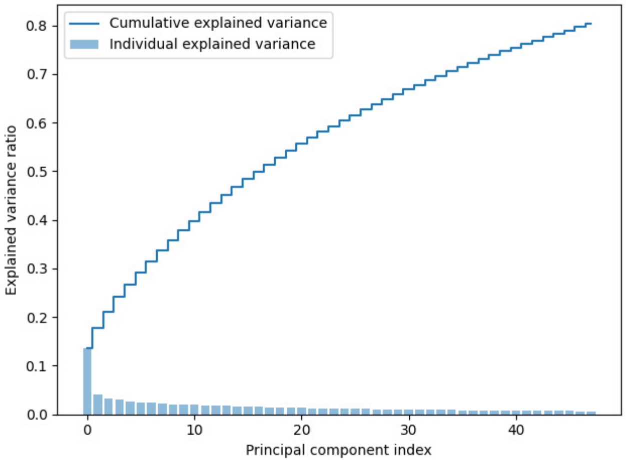


Figure 3: Cumulative Varience plot

4 Model Architecture

4.1 Model Performance Scores

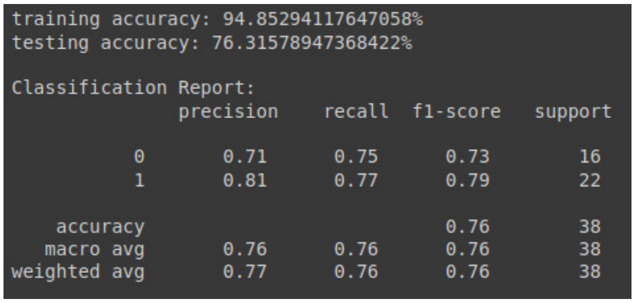
- The accuracy scores for the models are tabulated below :-

S. No	Classifier Used	Accuracy Score(%)
1	SVM	76.31
2	KNN	55
3	Decision Trees	53
4	Xtreme Gradient Boosting	71
5	Random Forests	65.78

- SVM and XGB models have better accuracy than other models.

4.2 Model training and Evaluation

- Here are the snapshots of the evaluation metrics of each model :-



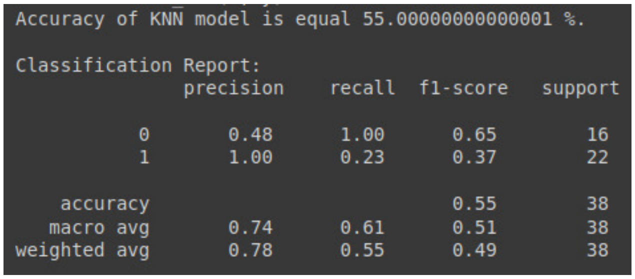
```
training accuracy: 94.85294117647058%
testing accuracy: 76.31578947368422%

Classification Report:
              precision    recall  f1-score   support

           0       0.71      0.75      0.73        16
           1       0.81      0.77      0.79        22

    accuracy                           0.76        38
   macro avg       0.76      0.76      0.76        38
weighted avg       0.77      0.76      0.76        38
```

Figure 4: SVM



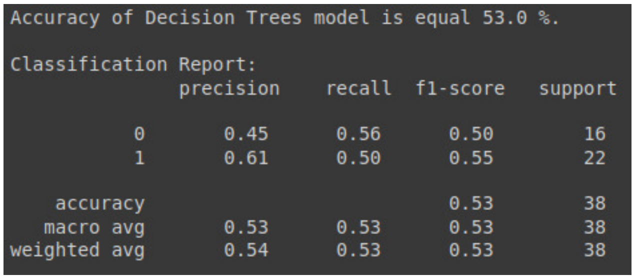
```
Accuracy of KNN model is equal 55.00000000000001 %.

Classification Report:
              precision    recall  f1-score   support

           0       0.48      1.00      0.65        16
           1       1.00      0.23      0.37        22

    accuracy                           0.55        38
   macro avg       0.74      0.61      0.51        38
weighted avg       0.78      0.55      0.49        38
```

Figure 5: KNN



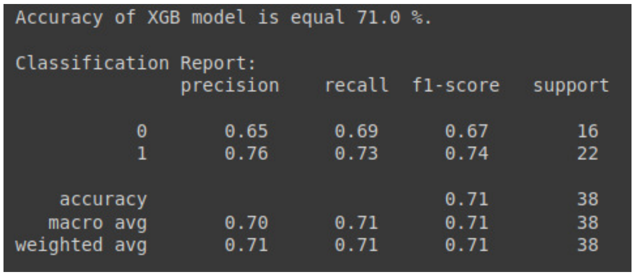
```
Accuracy of Decision Trees model is equal 53.0 %.

Classification Report:
              precision    recall  f1-score   support

           0       0.45      0.56      0.50        16
           1       0.61      0.50      0.55        22

    accuracy                           0.53        38
   macro avg       0.53      0.53      0.53        38
weighted avg       0.54      0.53      0.53        38
```

Figure 6: Decision Trees



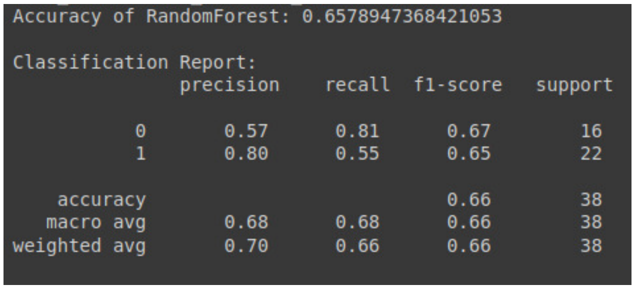
```
Accuracy of XGB model is equal 71.0 %.

Classification Report:
              precision    recall  f1-score   support

           0       0.65      0.69      0.67        16
           1       0.76      0.73      0.74        22

    accuracy                           0.71        38
   macro avg       0.70      0.71      0.71        38
weighted avg       0.71      0.71      0.71        38
```

Figure 7: XGB



```
Accuracy of RandomForest: 0.6578947368421053

Classification Report:
              precision    recall  f1-score   support

           0       0.57      0.81      0.67        16
           1       0.80      0.55      0.65        22

    accuracy                           0.66        38
   macro avg       0.68      0.68      0.66        38
weighted avg       0.70      0.66      0.66        38
```

Figure 8: Random Forests