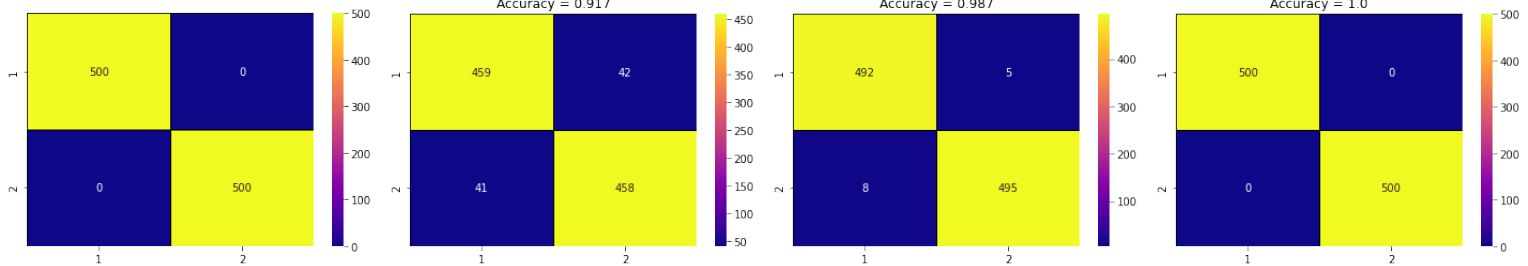# CS5691: Assignment 4

Akshat Meena (CS19B052), Shubham Patel (ME19B170)

A. Synthetic Dataset

1. Confusion Matrix

- We get 100% accuracy for KNN and ANN(using 500+ nodes in 2 hidden layer), we get least accuracy for LR.
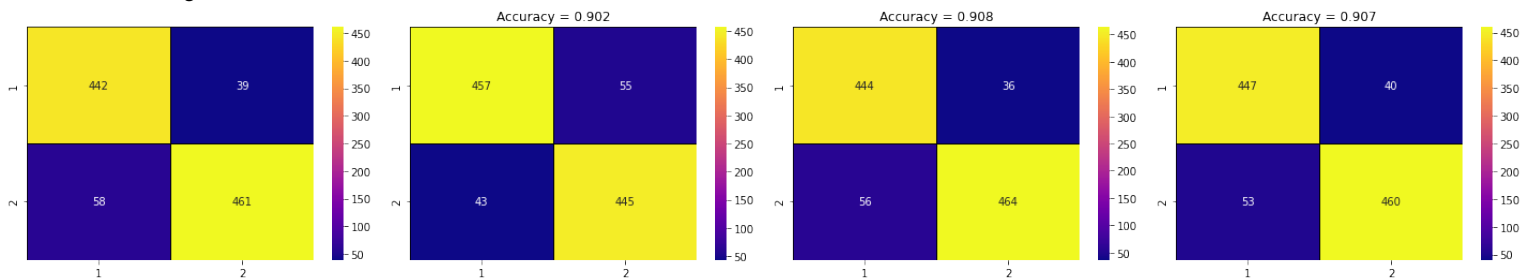
Figure 1: Confusion Matrix for Classification of synthetic data using KNN(k=10), Logistic Regression, SVM, ANN respectively

2. PCA

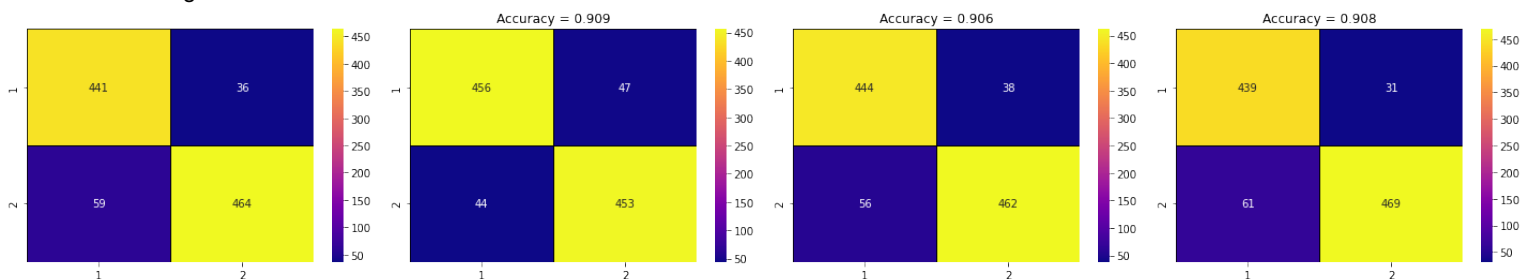- After PCA the accuracy is reduced to around 90-91% for all classifications.

Figure 2: Confusion Matrix for Classification of synthetic data using KNN(k=10), Logistic Regression, SVM, ANN respectively after reducing features to 1 using PCA


3. LDA

- We can observe that LDA is slightly better in some classification than the PCA for the same number of reduced dimensions

Figure 3: Confusion Matrix for Classification of synthetic data using KNN(k=10), Logistic Regression, SVM, ANN respectively after reducing features to 1 using LDA
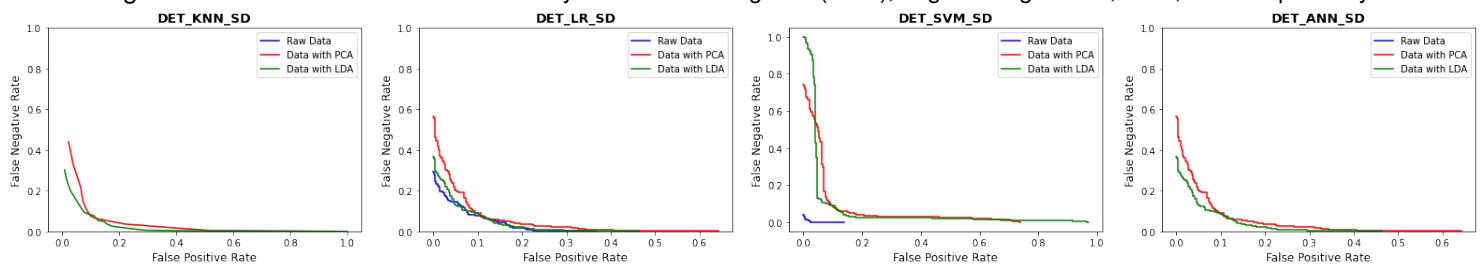

4. ROC

Figure 4: ROC curve for Classification of synthetic data using KNN(k=10), Logistic Regression, SVM, ANN respectively and comparing each classification.
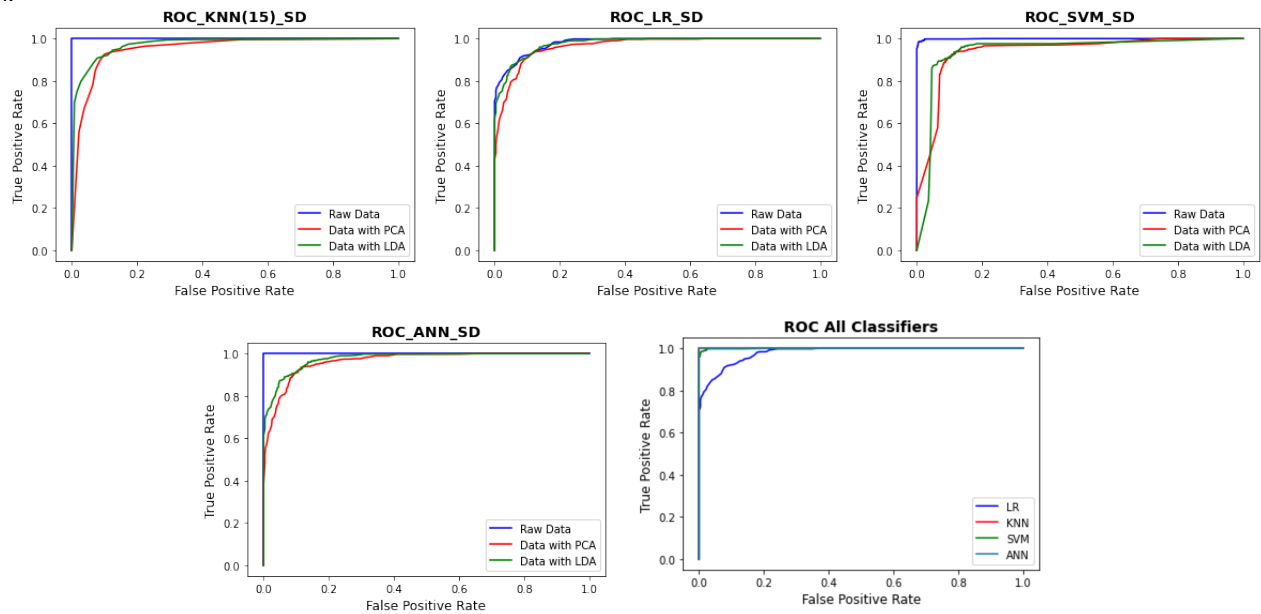
5. **DET**

Figure 5: DET curve for Classification of synthetic data using KNN(k=10), Logistic Regression, SVM, ANN respectively
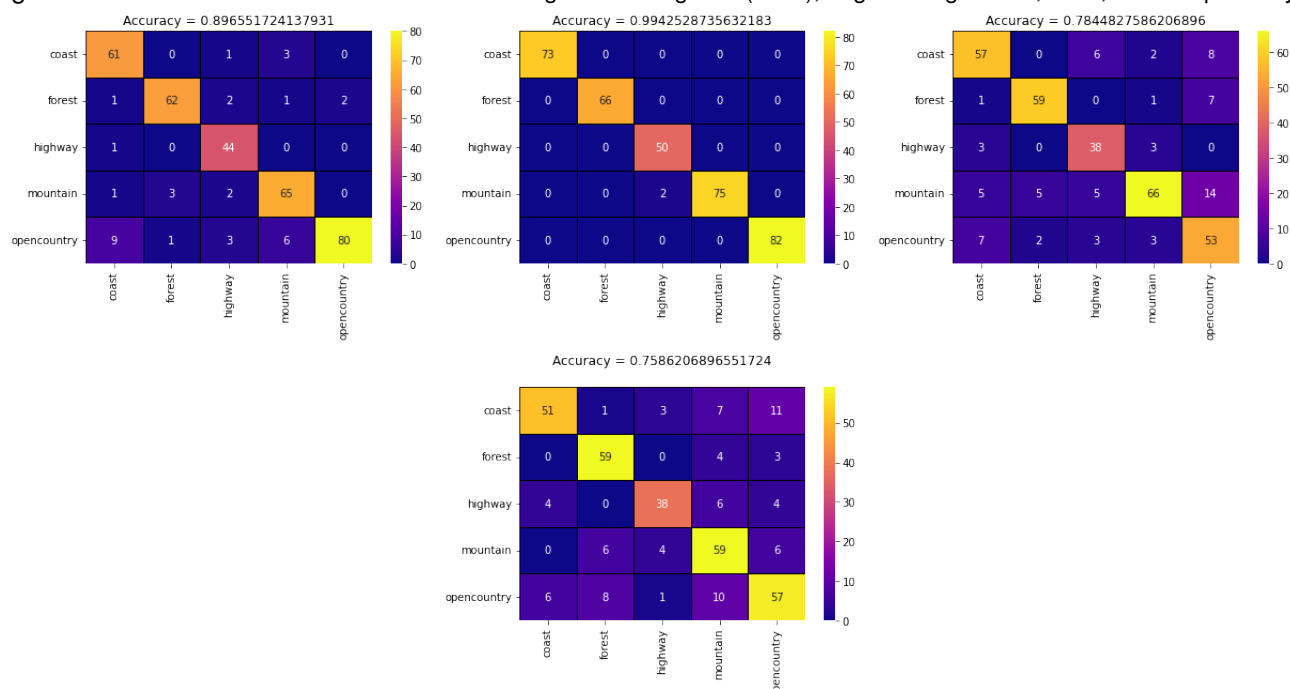

B. Image Dataset

1. **Confusion Matrix**

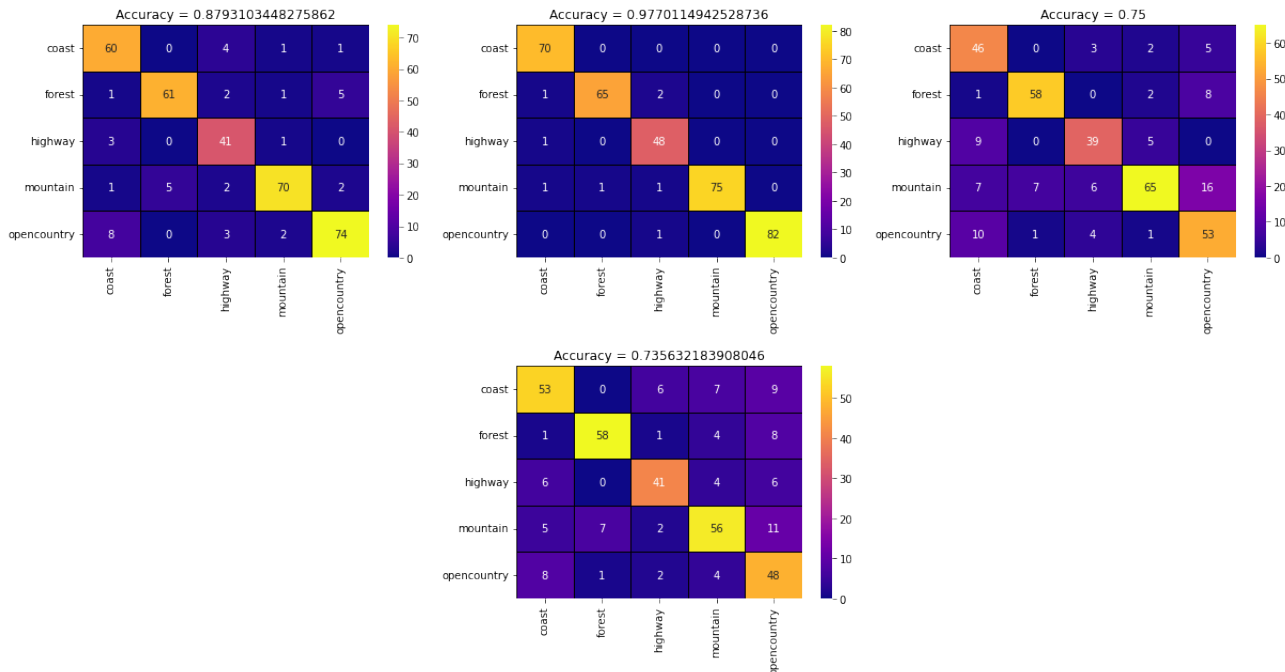   • Accuracy for each model LR>KNN>SVM>ANN

Figure 6: Confusion Matrix for Classification of image data using KNN(k=10), Logistic Regression, SVM, ANN respectively

2. **PCA**

- Even after reducing that dimension from 828 to 25 using PCA we can see that large percentage of data is preserved and we get pretty good accuracy.
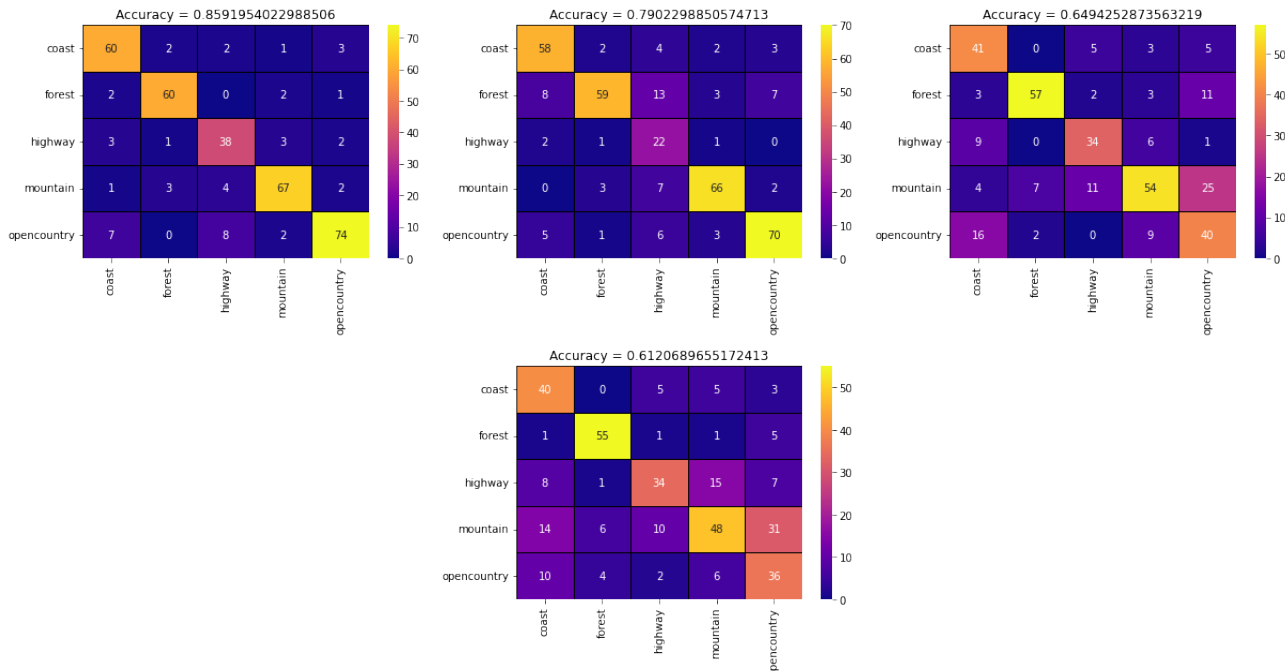
Figure 7: Confusion Matrix for Classification of image data using KNN(k=10), Logistic Regression, SVM, ANN respectively after reducing features to 25 using PCA


3. **LDA**

- We see that LDA is poorer than PCA for same number of components on image data. Also we observed that there is not much change in the accuracy for taking some more number of components than 25 too.

Figure 8: Confusion Matrix for Classification of image data using KNN(k=10), Logistic Regression, SVM, ANN respectively after reducing features to 25 using LDA
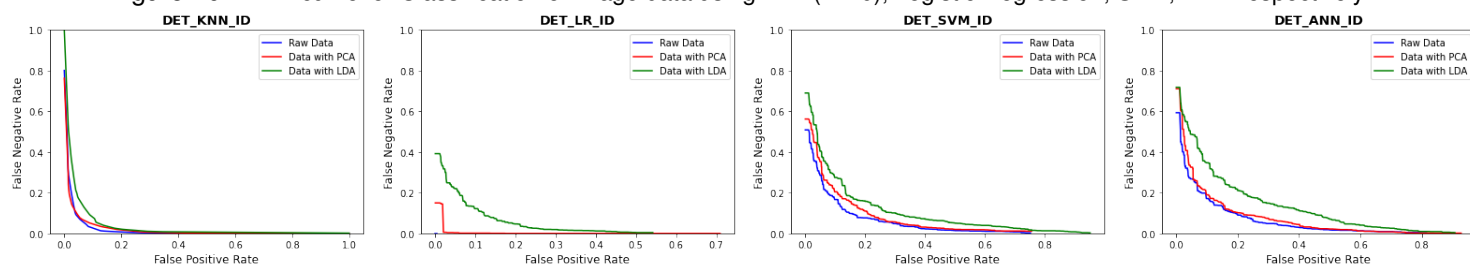

4. **ROC**

Figure 9: ROC curve for Classification of image data using KNN(k=10), Logistic Regression, SVM, ANN respectively and comparing each classification.
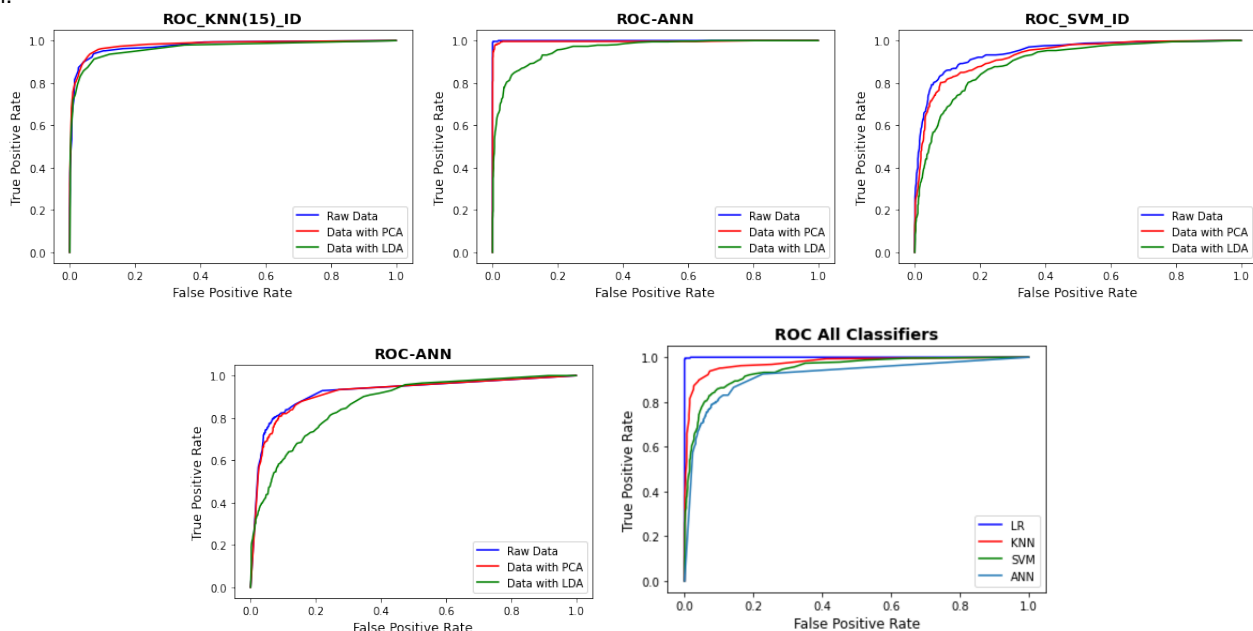
5. DET

Figure 10: DET curve for Classification of image data using KNN(k=10), Logistic Regression, SVM, ANN respectively
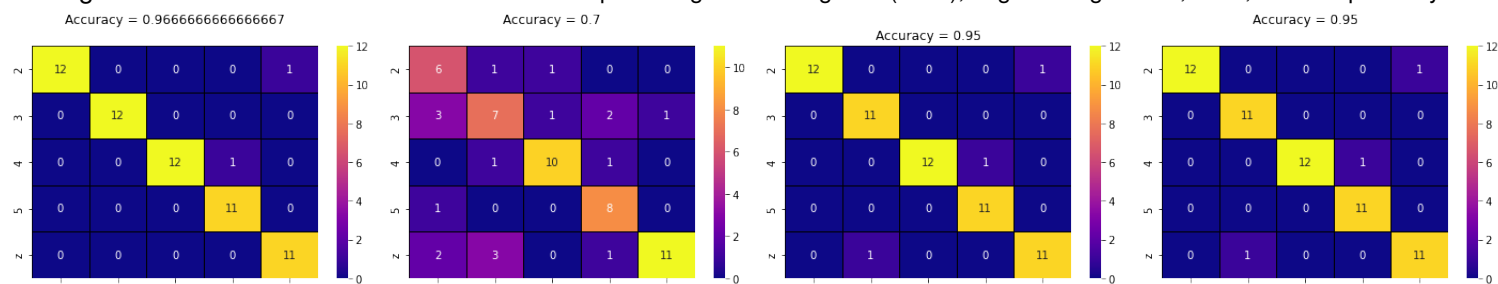

C. Spoken Digit

1. Confusion Matrix

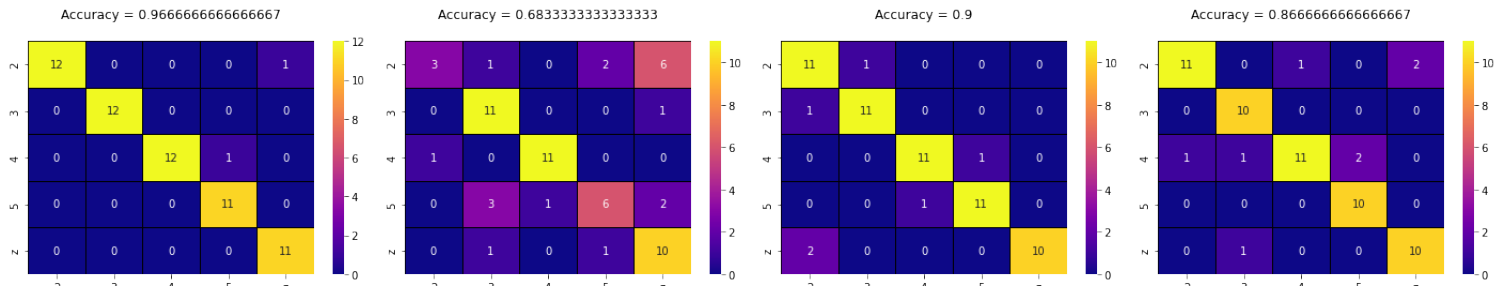- On spoken digit data all classifier except LR give good accuracy

Figure 11: Confusion Matrix for Classification of spoken digit data using KNN(k=10), Logistic Regression, SVM, ANN respectively

2. PCA

- After reducing dimensions to 15 we observe that we still get good accuracy for KNN whereas for SVM and ANN it reduces significantly relatively.
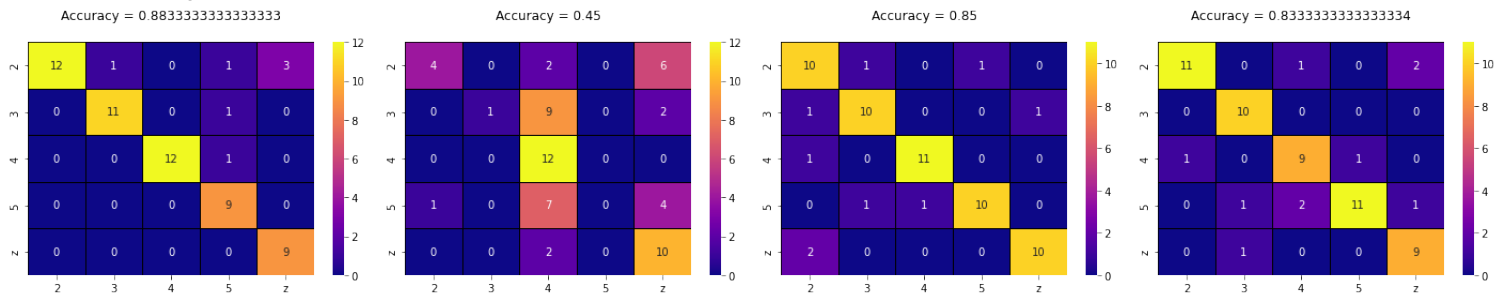
Figure 12: Confusion Matrix for Classification of spoken digit data using KNN(k=10), Logistic Regression, SVM, ANN respectively after reducing features to 15 using PCA
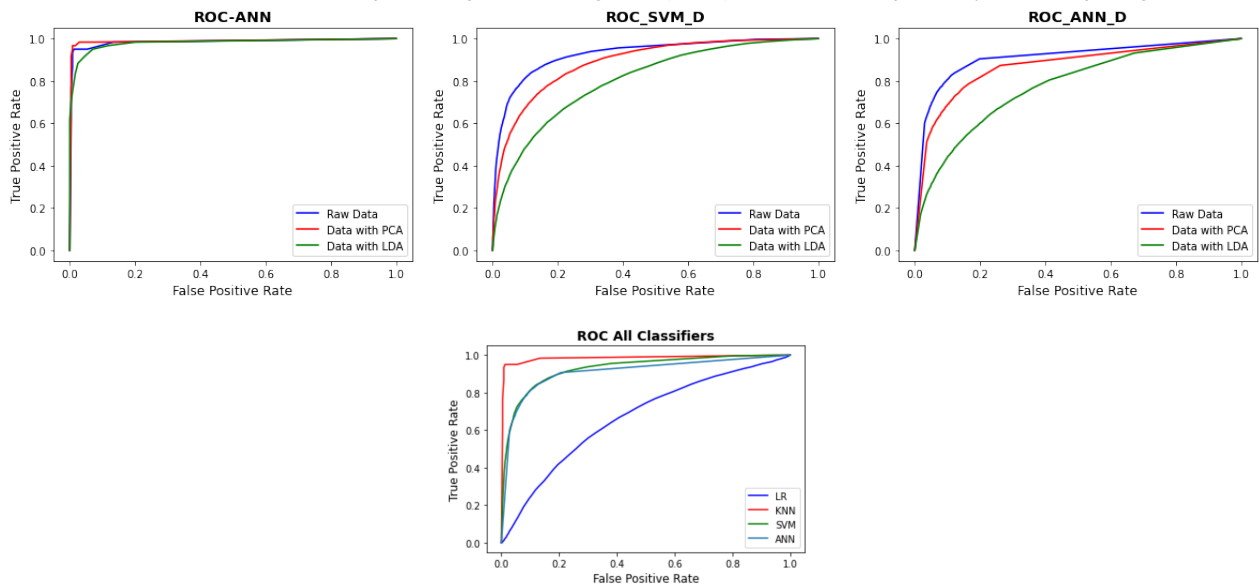

3. **LDA**

   • Comparative to PCA we get poorer accuracy using LDA

Figure 13: Confusion Matrix for Classification of spoken digit data using KNN(k=10), Logistic Regression, SVM, ANN respectively after reducing features to 15 using PCA
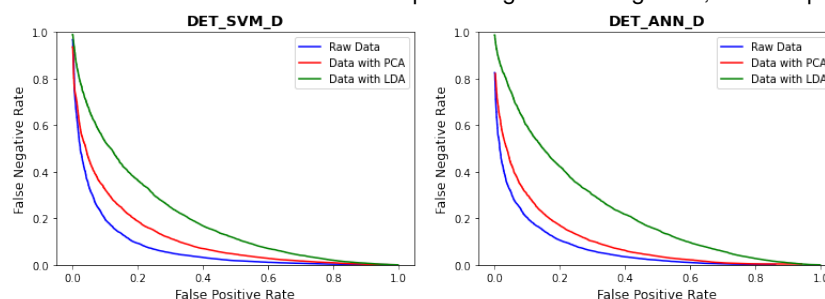

4. **ROC**

Figure 14: ROC curve for Classification of spoken digit data using KNN(k=10) SVM, ANN respectively and comparing each classification.
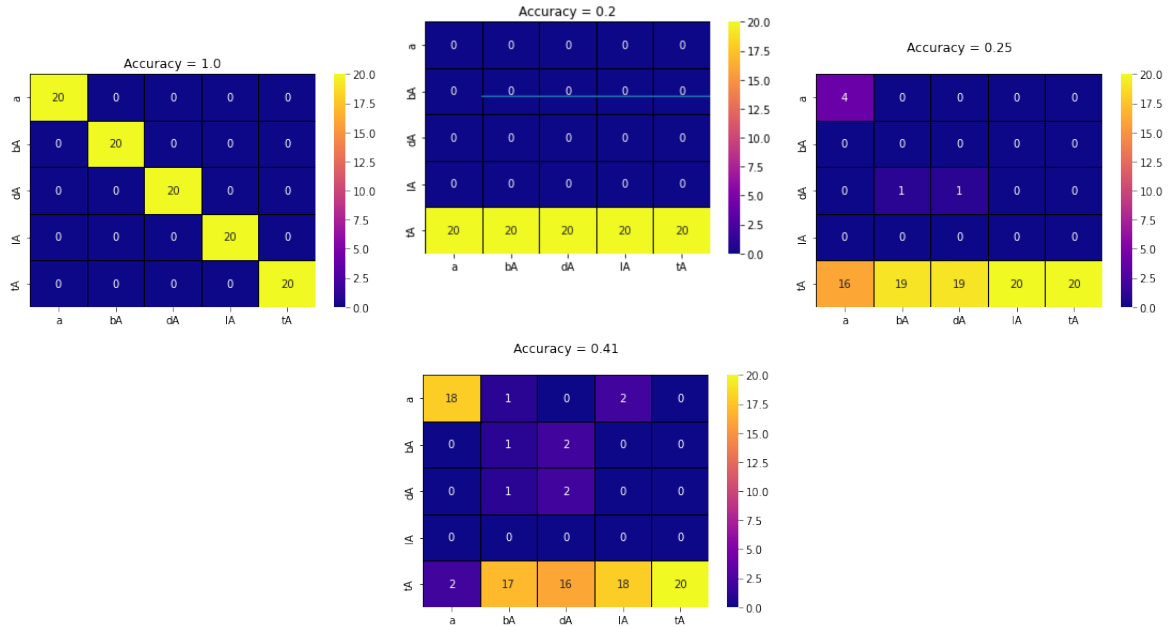

5. **DET**

Figure 15: DET curve for Classification of spoken digit data using SVM, ANN respectively
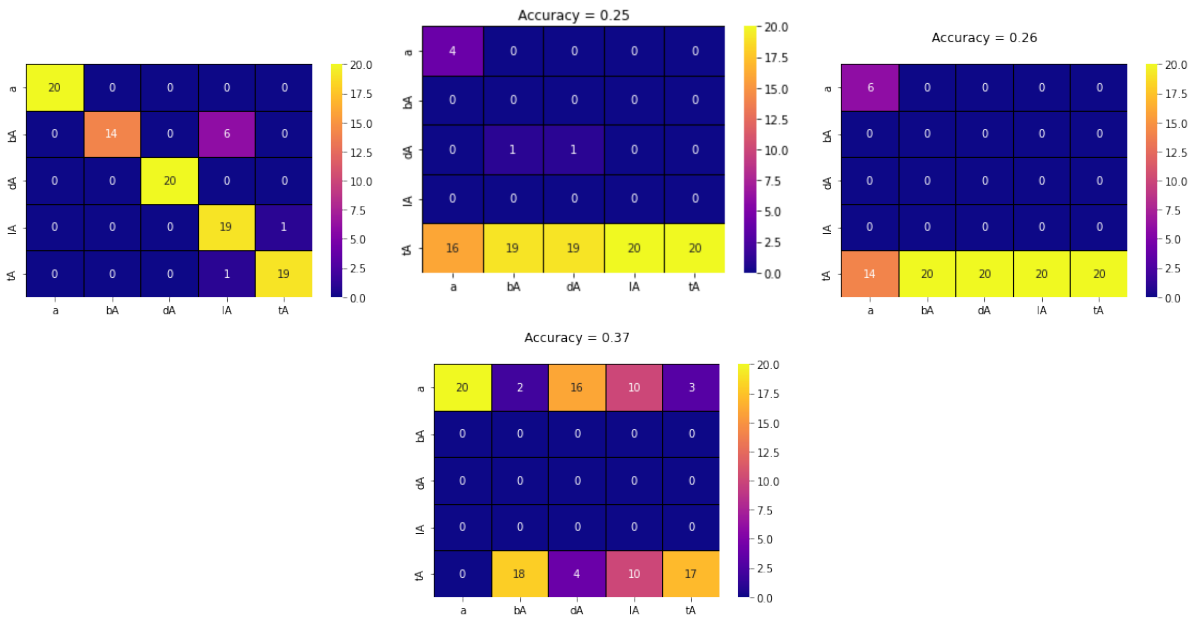
D. Online Handwritting

1. **Confusion Matrix**

Figure 16: Confusion Matrix for Classification of online handwritting data using KNN(k=10), Logistic Regression, SVM, ANN respectively
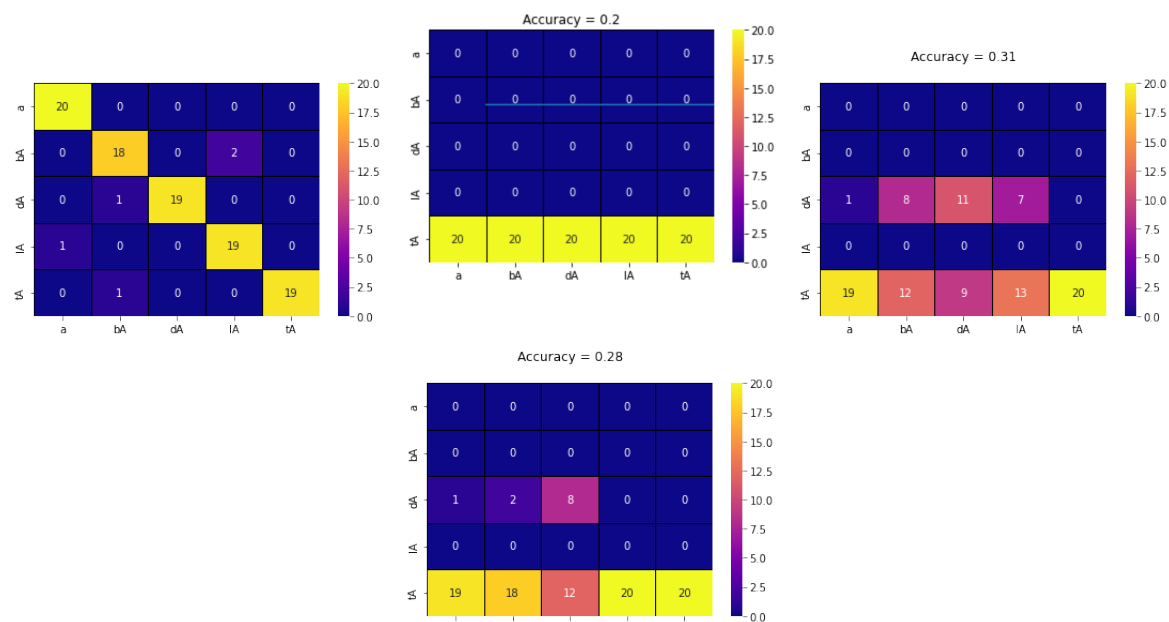

2. **PCA**

Figure 17: Confusion Matrix for Classification of online handwritting data using KNN(k=10), Logistic Regression, SVM, ANN respectively after reducing features to 1 using PCA

3. **LDA**

Figure 18: Confusion Matrix for Classification of online handwriting data using KNN(k=10), Logistic Regression, SVM, ANN respectively after reducing features to 1 using PCA
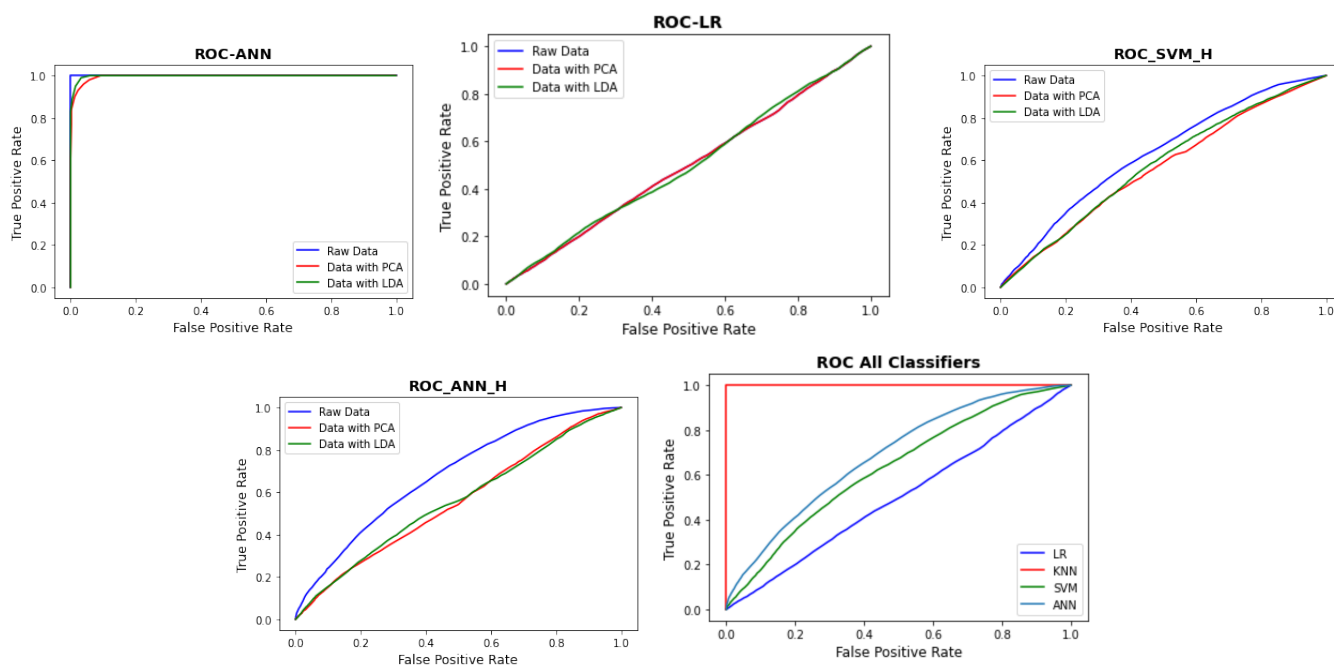

4. ROC

Figure 19: ROC curve for Classification of online handwriting data using KNN(k=10), Logistic Regression, SVM, ANN respectively and comparing each classification.

5. DET

Figure 20: DET curve for Classification of online handwriting data using SVM, ANN respectively