# SMAI Assignment 2 Report

Tanuj Garg
20171106
International Institute of Information Technology, Hyderabad
tanuj.garg@students.iiit.ac.in

## I. Introduction

This report contains data analysis and results of various classification techniques used on 3 different problems involving datasets containing face images of humans. Different feature representations for the images have been tried. In the end, a new problem statement is described with its dataset, data analysis and classification results.

## II. Dimensionality Reduction

### A. Eigenfaces

Eigenface is a method that is useful for face recognition and detection by determining the variance of faces in a collection of face images and use those variances to encode and decode a face in a machine learning way without the full information reducing computation and space complexity.

### B. Minimum eigen vectors required

To find the minimum eigen vectors required to satisfactorily reconstruct dataset, eigen spectrum of the datasets are observed. Number of non-zero eigen values are approximated and those many eigen vectors are taken.
From Figure 1, 2 and 3, it can be observed that most of the eigen values are negligible. For each dataset, minimum eigen vectors required are listed

1) IMFDB - 80
2) IIIT-CFW - 130
3) Yale - 20

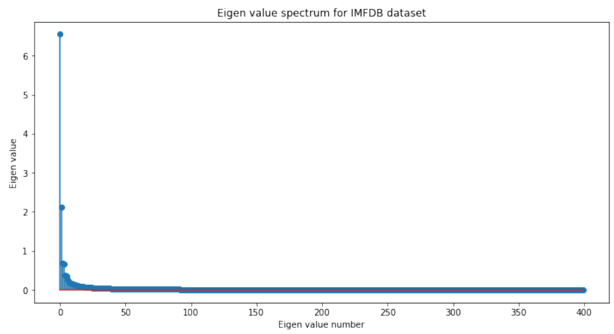


Fig. 1. Eigen Value Spectrum for IMFDB Dataset

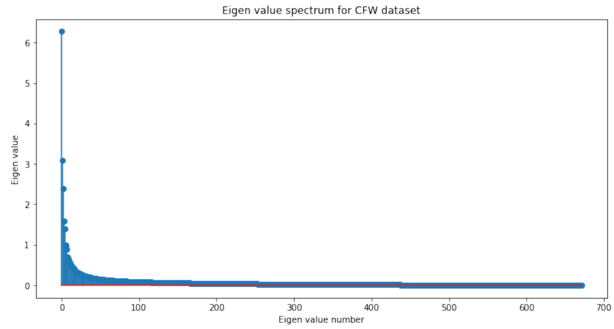


Fig. 2. Eigen Value Spectrum for IIIT-CFW Dataset

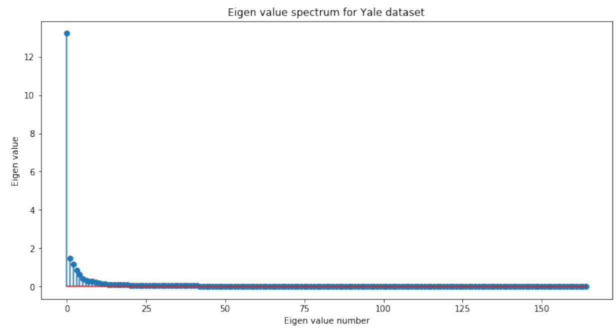


Fig. 3. Eigen Value Spectrum for Yale Dataset

### C. Entity/Dataset difficult to represent compactly

IIIT-CFW Dataset is most difficult to represent compactly. This dataset has high variance in images due to which it is hard to compactly represent them. This is because of their shapes of faces, color distribution, sharp edges and that cartoons have more distinct features than human faces (which have common features like nose, eyes etc.). It can also be seen from its eigen value spectrum that more eigen vectors are required to represent this dataset. Also, reconstruction error is highest for this dataset (0.124 for this and 0.05,0.098 for others). Some examples are in Figure 4.

Identities which are hard to represent compactly are decided by reconstruction error. This can also be similarly explained by looking at some of the examples of these.

1) IMFDB - Shilpa Shetty (Figure 5)
2) IIIT-CFW - Narendra Modi (Figure 6)
3) Yale - Class 7 (Figure 7)

Fig. 4. Examples from IIIT-CFW dataset

Fig. 5. Examples of Shilpa Shetty from IMFDB dataset

Fig. 6. Examples of Narendra Modi from IIIT-CFW dataset

Fig. 7. Examples of class 7 from Yale dataset

## III. CLASSIFICATION

### A. Comparative study of different classification techniques

For each algorithm, different parameters have been tried and only the best results are displayed

| IMFDB Dataset | | | | |
|---|---|---|---|---|
| Algorithm | Validation Accuracy | Precision | Recall | F1 score |
| Logistic Regression | 82.5% | 0.83 | 0.84 | 0.83 |
| SVM | 83.75% | 0.85 | 0.84 | 0.83 |
| Decision Tree | 53.75% | 0.54 | 0.55 | 0.53 |
| MLP | 87.5% | 0.87 | 0.88 | 0.87 |

| IIIT-CFW Dataset | | | | |
|---|---|---|---|---|
| Algorithm | Validation Accuracy | Precision | Recall | F1 score |
| Logistic Regression | 62.96% | 0.59 | 0.6 | 0.58 |
| SVM | 68.15% | 0.65 | 0.66 | 0.64 |
| Decision Tree | 34.07% | 0.32 | 0.31 | 0.31 |
| MLP | 61.48% | 0.57 | 0.56 | 0.55 |

| Yale Dataset | | | | |
|---|---|---|---|---|
| Algorithm | Validation Accuracy | Precision | Recall | F1 score |
| Logistic Regression | 93.94% | 0.91 | 0.89 | 0.9 |
| SVM | 90.91% | 0.89 | 0.85 | 0.86 |
| Decision Tree | 75.76% | 0.74 | 0.69 | 0.68 |
| MLP | 100% | 1.0 | 1.0 | 1.0 |

### B. Classification on reduced dimension space

Many different combinations of feature representations were tried and different classifiers were used and only the best results are displayed

| IMFDB | | | | |
|---|---|---|---|---|
| Method | Reduced Space | Error | Accuracy | F1 score |
| (LDA+PCA)+SVM | 87 | 0.0031 | 77.5% | 0.78 |
| ResNet+MLP | 2048 | 0.0 | 95% | 0.95 |
| (VGG+ResNet)+LR | 6144 | 0.0 | 100% | 1.00 |
| (VGG+ResNet)+SVM | 6144 | 0.0 | 98.75% | 0.98 |
| All features+MLP | 6318 | 0.0 | 100% | 1.00 |
| (KLDA+ResNet)+LR | 2055 | 0.0 | 97.5% | 0.97 |
| (KPCA+LDA)+SVM | 87 | 0.0094 | 68.75% | 0.71 |
| KLDA+LR | 7 | 0.2781 | 62.5% | 0.60 |

| IIIT-CFW Dataset | | | | |
|---|---|---|---|---|
| Method | Reduced Space | Error | Accuracy | F1 score |
| (LDA+PCA)+SVM | 127 | 0.0447 | 50.37% | 0.48 |
| ResNet+MLP | 2048 | 0.0 | 97.04% | 0.96 |
| (VGG+ResNet)+LR | 6144 | 0.0 | 99.26% | 0.99 |
| (VGG+ResNet)+SVM | 6144 | 0.0112 | 100% | 1.00 |
| All features+MLP | 6398 | 0.0 | 100% | 1.00 |
| (KLDA+ResNet)+LR | 2055 | 0.0 | 94.07% | 0.93 |
| (KPCA+LDA)+SVM | 127 | 0.0354 | 34.07% | 0.31 |
| KLDA+LR | 7 | 0.2551 | 29.63% | 0.26 |

| Yale Dataset | | | | |
|---|---|---|---|---|
| Method | Reduced Space | Error | Accuracy | F1 score |
| (LDA+PCA)+SVM | 44 | 0.0 | 96.97% | 0.92 |
| ResNet+MLP | 2048 | 0.0 | 96.97% | 0.96 |
| (VGG+ResNet)+LR | 6144 | 0.0 | 100% | 1.00 |
| (VGG+ResNet)+SVM | 6144 | 0.0 | 100% | 1.00 |
| All features+MLP | 6232 | 0.0 | 100% | 1.00 |
| (KLDA+ResNet)+LR | 2062 | 0.0 | 100% | 1.00 |
| (KPCA+LDA)+SVM | 44 | 0.0 | 100% | 1.00 |
| KLDA+LR | 14 | 0.0 | 93.94% | 0.95 |

Some observations -

1) On concatenating all features, maximum accuracy is obtained because most information is captured but space required is also maximum
2) Neural Network representations VGG and ResNet are performing way better than other representations
3) Accuracies are best in Yale dataset, thus implying that there is enough separability in dataset

## IV. t-SNE BASED VISUALISATION OF FACES

After doing t-SNE based visualisation for each dataset separately, clustering is not observed since it is an unsupervised algorithm which does not use information from the class labels. But when we do this on combined dataset, we see points corresponding to same dataset are closer (Figure 8 for 2D and 9 for 3D). If we apply lda and then do tsne then, clustering is observed because lda is supervised algorithm (Figure 10).

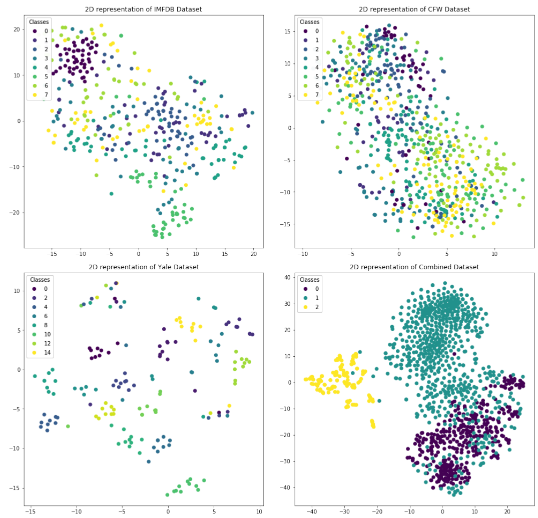


Fig. 8. 2D t-SNE representation

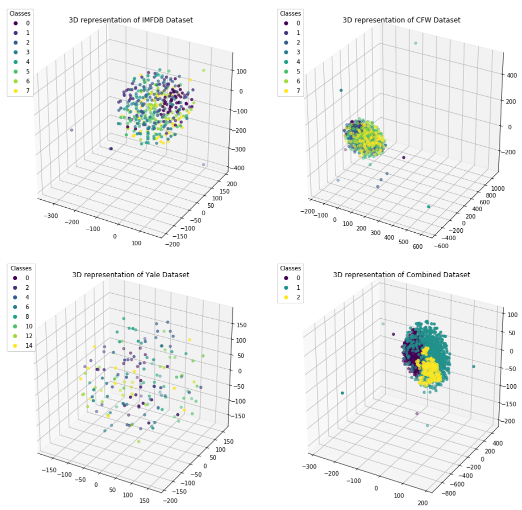


Fig. 9. 3D t-SNE representation

## V. K-NEAREST NEIGHBOR

### A. Problem Formulation

Given an image, its distance from all images in dataset can be calculated using any distance measure (it can be done pixel by pixel or by using any other representation). Then, current image can be classified using the majority

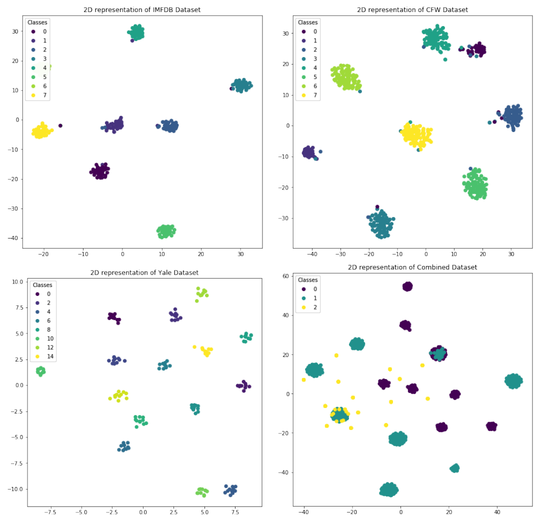


Fig. 10. 2D t-SNE representation after lda

voting method on k nearest images and if it is equal to classID in input, then response is ”Yes”, else ”No”.

### B. Performance Analysis

For the task of face verification, precision metric is most appropriate since we want to minimize false positives (we don’t want to give access to some intruder by mistake). Accuracy can also be used for this.

Performance details of KNN for different datsets, feature representations and k values are summarised in following tables.

| IMFDB Dataset | | | | | |
|---|---|---|---|---|---|
| Method | k | Reduced Space | Error | Accuracy | Precision |
| Raw Pixels | 3 | 3072 | 0.23 | 60% | 0.65 |
| LDA+PCA | 3 | 87 | 0.0344 | 80% | 0.84 |
| ResNet | 5 | 2048 | 0.0281 | 96.25% | 0.96 |
| VGG+ResNet | 5 | 6144 | 0.0125 | 98.75% | 0.99 |
| All features | 3 | 6318 | 0.0062 | 98.75% | 0.99 |
| KLDA+ResNet | 5 | 2055 | 0.0156 | 97.5% | 0.97 |
| KPCA+LDA | 5 | 87 | 0.025 | 80% | 0.85 |
| KLDA | 3 | 7 | 0.2406 | 50% | 0.47 |

| IIIT-CFW Dataset | | | | | |
|---|---|---|---|---|---|
| Method | k | Reduced Space | Error | Accuracy | Precision |
| Raw Pixels | 3 | 3072 | 0.52 | 31.85% | 0.55 |
| LDA+PCA | 3 | 87 | 0.1359 | 44.44% | 0.50 |
| ResNet | 5 | 2048 | 0.0205 | 97.78% | 0.97 |
| VGG+ResNet | 5 | 6144 | 0.0186 | 99.26% | 0.99 |
| All features | 3 | 6398 | 0.0093 | 83.7% | 0.82 |
| KLDA+ResNet | 5 | 2055 | 0.0223 | 82.96% | 0.83 |
| KPCA+LDA | 5 | 87 | 0.0391 | 39.26% | 0.40 |
| KLDA | 3 | 7 | 0.216 | 25.19% | 0.26 |

| Yale Dataset | | | | | |
|---|---|---|---|---|---|
| Method | k | Reduced Space | Error | Accuracy | Precision |
| Raw Pixels | 3 | 3072 | 0.09 | 90.91% | 0.94 |
| LDA+PCA | 3 | 87 | 0.0833 | 93.94% | 0.95 |
| ResNet | 5 | 2048 | 0.0076 | 100% | 1.00 |
| VGG+ResNet | 5 | 6144 | 0.0227 | 96.97% | 0.97 |
| All features | 3 | 6318 | 0.0 | 100% | 1.00 |
| KLDA+ResNet | 5 | 2055 | 0.0 | 100% | 1.00 |
| KPCA+LDA | 5 | 87 | 0.0 | 96.97% | 0.93 |
| KLDA | 3 | 7 | 0.0152 | 96.97% | 0.93 |

## VI. Extension/Application

### A. Problem Statement

Given a facial image, the problem is to decide whether it is an image of a cartoon or of a real person. This problem is non-trivial because a large amount of dataset is needed to train this model. Also, following features of cartoons makes it more difficult

1) Strong Colours
2) Sharp edges (unlike faces)
3) Varying Shapes (of faces and other features of the face)

### B. Applications

1) There are certain online platforms, where people are asked to upload their personal picture as Display Picture, but they often use non-real images. This classifier can detect that and ask them to upload a real picture
2) Since, in general cartoons have more features than normal face images (because of violation of shape, sharp color etc). We can separate them from rest of dataset using this classifier and do dimensionality reduction on them separately with more number of features.

### C. Dataset

Dataset for this problem is formed by combining the above 3 datasets. IMFDB Dataset (containing 400 images) and Yale Dataset (containing 165 images) are real pictures of humans considered as class 0 while IIIT-CFW Dataset (containing 672 images) are cartoon pictures considered as class 1. Since there are 565 images of class 0 and 672 images of class 1, there is no class imbalance.
3D T-SNE visualisation of dataset is given in Figure 11

### D. Pipeline

First the dataset was split into 2 parts, 80% for training and rest 20% for testing
Kernel Principle Component Analysis with 3 degree rbf kernel was used for feature representation of dataset reducing it to 80 dimensions per image.
Then, a Multi Layer Perceptron Classifier was used with input layer of size 80, 2 hidden layers of size 100 each and output layer of size 2. ReLU activation function was used for hidden layersn and adam optimization was done.

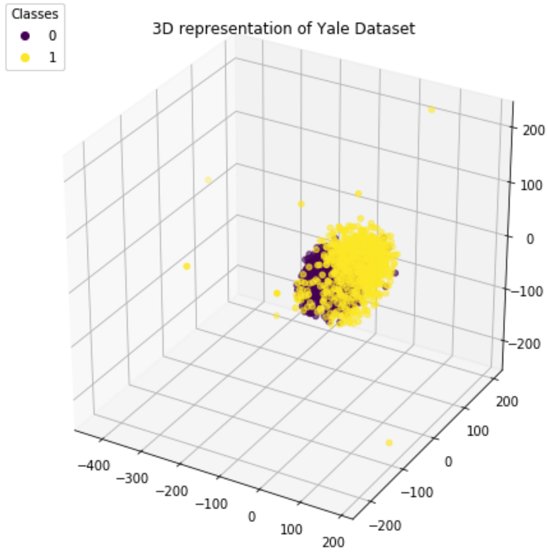


Fig. 11. 3D t-SNE representation

### E. Results

Accuracy on test data is 97.18%. Precision score is 1.0 and recall score is 0.95. F1 Score is 0.97
Confusion Matrix is shown in Figure 12

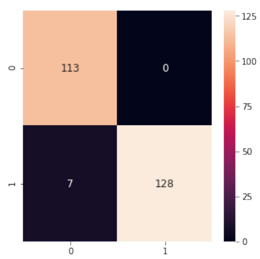


Fig. 12. Confusion Matrix

Results on k-fold validation on 5 folds are represented in Figure 13

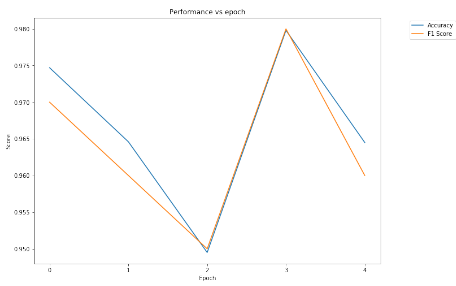


Fig. 13. Accuracy and F1 score vs epoch in 5-fold validation

Some examples of images which were correctly classified are in Figure 14, while incorrectly classified in Figure 15


Fig. 14. Correctly Classified Samples

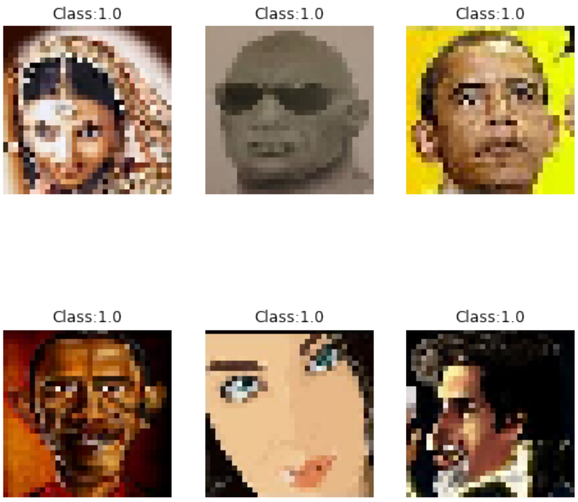


Fig. 15. Wrongly Classified Samples

All of the images which are wrongly classified were classified as real persons instead of cartoons. It is even difficult to do this with human eye, because some of the cartoons are too realistic (first 3 images). Also, since the dataset was limited, it could not capture different angles of face image like image 5.

## References

[1] Ming-Hsuan Yang. Face Recognition using Kernel Methods

[2] Yale Face Database,Url:https://vismod.media.mit.edu/vismod/classes/mas622-00/dataset

[3] Shankar Setty, Moula Husain, Parisa Beham, Jyothi Gudavalli, Menaka Kandasamy, Radhesyam Vaddi, Vidyagouri Hemadri, J C Karure, Raja Raju, Rajan, Vijay Kumar and C V Jawahar. ”Indian Movie Face Database: A Benchmark for Face Recognition Under Wide Variations”, NCVPRIPG-2013

[4] Mishra, A., Nandan Rai, S., Mishra, A. and Jawahar, C. V., ”IIIT-CFW: A Benchmark Database of Cartoon Faces in the Wild”, VASE ECCVW-2016