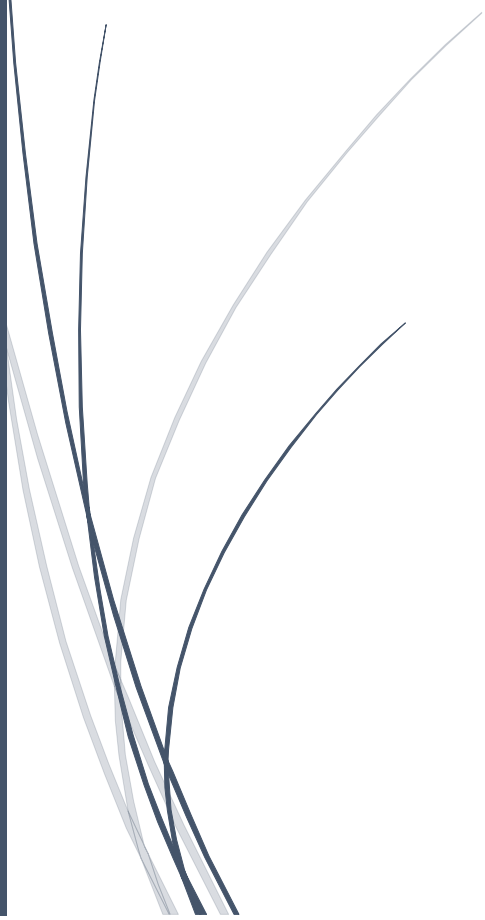
project

Face Recognition

Team Members

Mohamad Elsyed Mohamad Ahmed	**2305273**
Yasser Ali Mohamad Mohamad	**2305268**
Mohamad Ashraf Fathy Ahmed	**2305181**



Face Recognition

Project Overview:

The goal of this project is to implement and evaluate face recognition techniques using dimensionality reduction methods and simple classifiers. The assignment utilizes the ORL (AT&T) face dataset, which contains grayscale images of 40 individuals, each represented by 10 facial images. The primary objective is to identify the correct subject (person) from an input image by comparing it with a database of known faces.

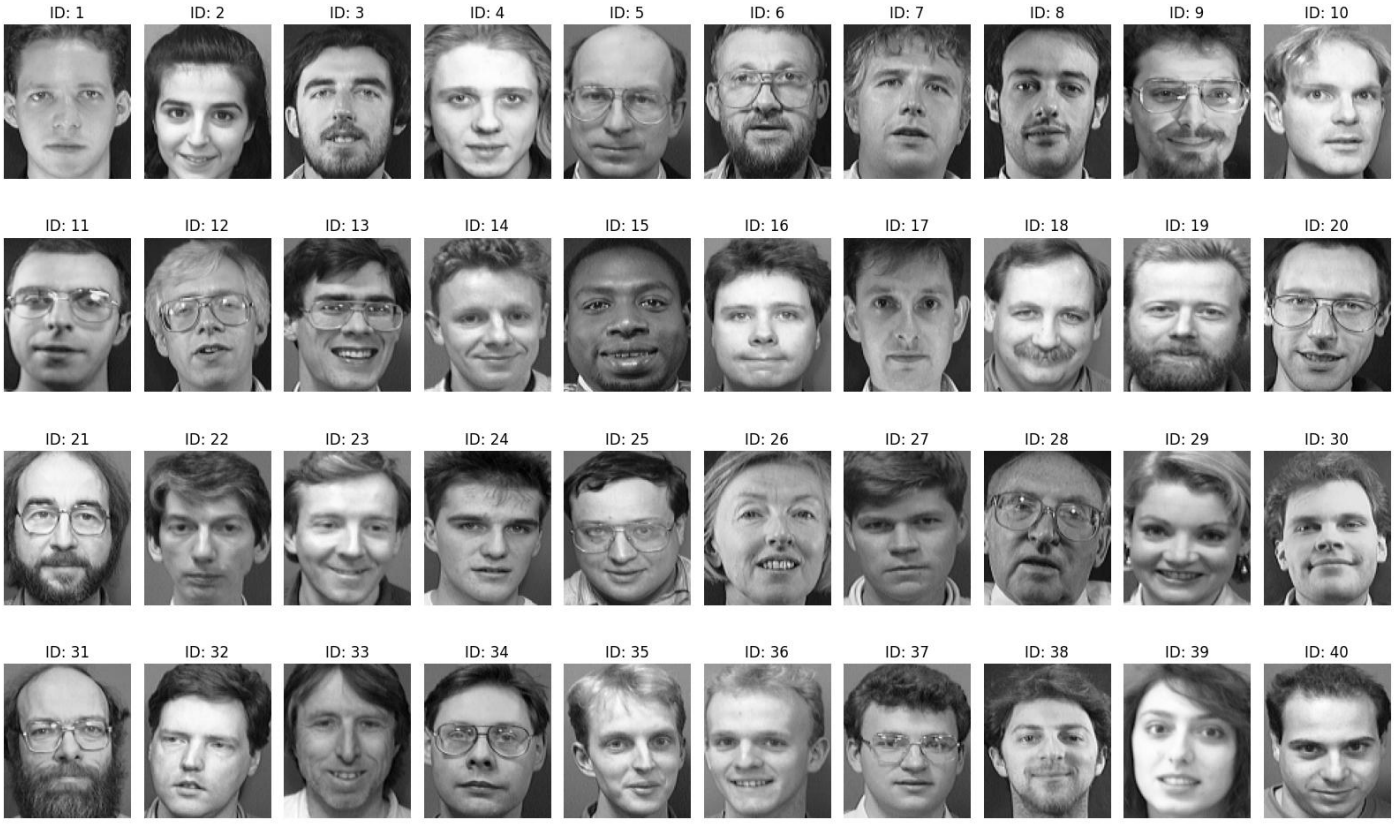
To achieve this, the project is divided into multiple stages:

- **Preprocessing the dataset and organizing it into suitable formats for analysis.**
- **Applying Principal Component Analysis (PCA) to reduce dimensionality while retaining most of the image variance.**
- **Using Linear Discriminant Analysis (LDA) for supervised dimensionality reduction by maximizing class separability.**
- **Employing the K-Nearest Neighbors (K-NN) algorithm for classification in the reduced feature space.**
- **Evaluating model accuracy and tuning the classifier with different parameters.**
- **Optionally, extending the task to differentiate between face and non-face images.**

This assignment not only demonstrates how classical machine learning techniques can be used for high-dimensional data like images but also provides insights into how preprocessing and dimensionality reduction affect classification performance.

1. Dataset Description

The dataset used in this project is the ORL Database of Faces, which contains 40 subjects, each with 10 grayscale images of size 92x112. The dataset was downloaded from Kaggle.



2. Data Preprocessing

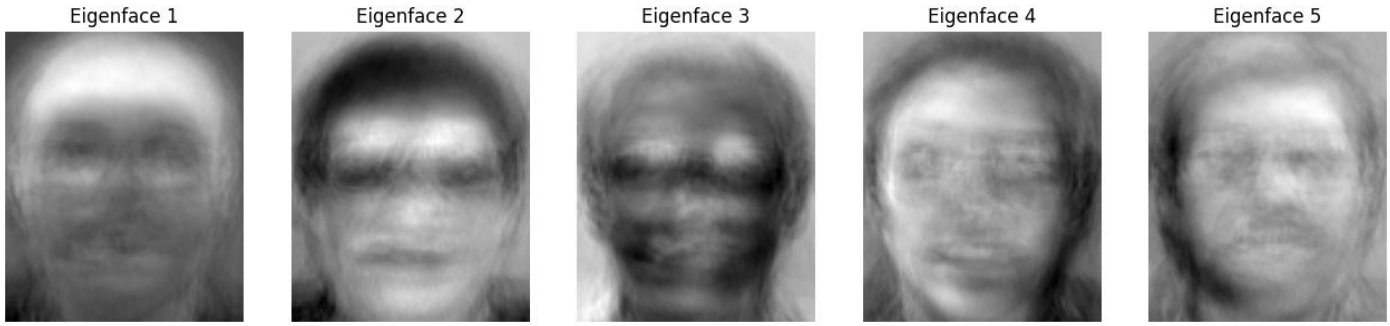
Each image was reshaped into a 1D vector of 10304 elements. A data matrix $D \in R^{400\times10304}$ was created, along with a label vector $y \in R^{400}$,with labels from 1 to 40.

3. Data Splitting

The dataset was split such that the odd-indexed images were used for training (5 per subject) and the even-indexed images were used for testing (5 per subject).

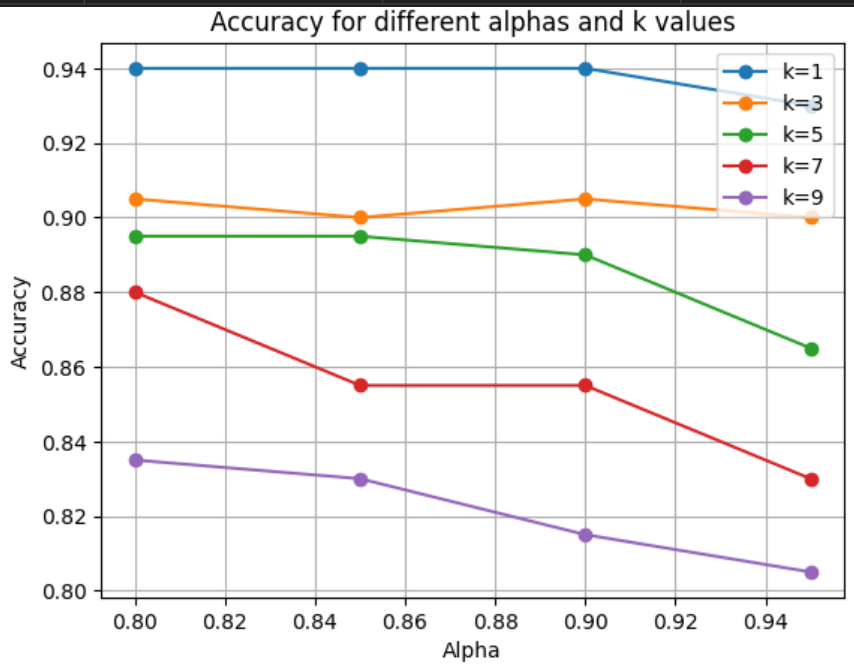
4. PCA Classification

We applied PCA with different values of α the proportion of variance retained. A 1-Nearest Neighbor classifier was used for evaluation.



Accuracy for different alphas and k values

	1	3	5	7	9
0.8	0.94	0.905	0.895	0.88	0.835
0.85	0.94	0.9	0.895	0.855	0.83
0.9	0.94	0.905	0.89	0.855	0.815
0.95	0.93	0.9	0.865	0.83	0.805

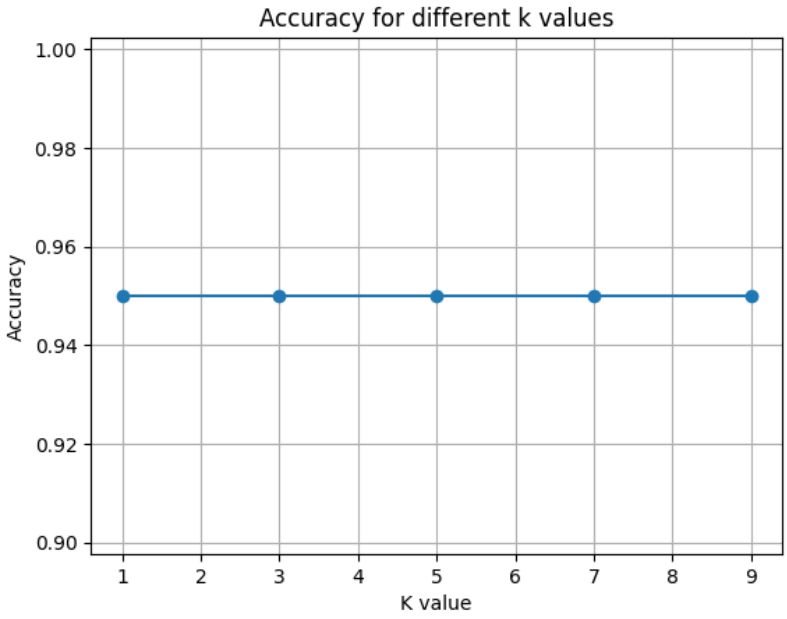


PCA Accuracy: 0.955

5. LDA Classification

Multiclass LDA was applied using 39 dominant eigenvectors. A 1-NN classifier was used to evaluate performance.

LDA Accuracy: 0.95



6. Comparison Between PCA & LDA

Accuracy of PCA VS LDA with respect to K (number of nearest neighbors)

	1	3	5	7	9
PCA α = 0.8	94.0%	90.5%	89.5%	88.0%	83.5%
PCA α = 0.85	94.0%	90.0%	89.5%	85.5%	83.0%
PCA α = 0.9	94.0%	90.5%	89.0%	85.5%	81.5%
PCA α = 0.95	93.0%	90.0%	86.5%	83.0%	80.5%
LDA	95.0%	95.0%	95.0%	95.0%	95.0%

7. Bonus Section

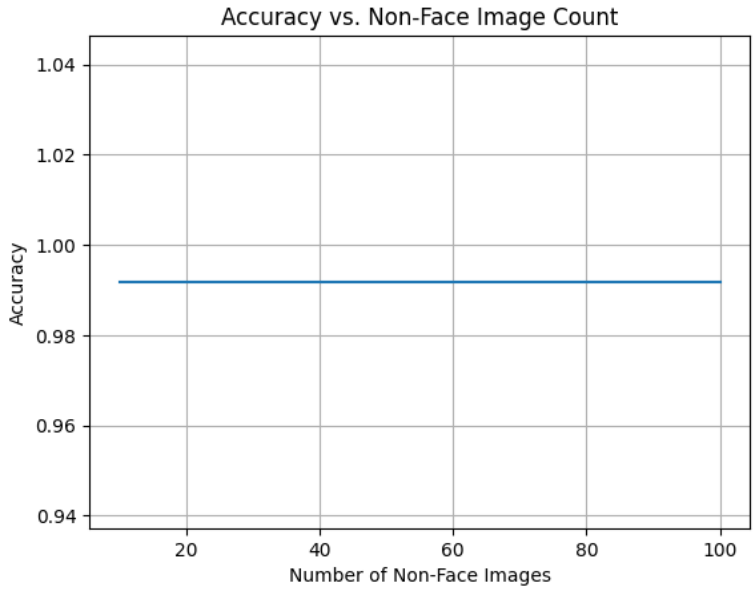
a. Faces vs Non-Faces Classification

```
Accuracy: 0.9918032786885246
              precision    recall  f1-score   support

        face       0.99      1.00      1.00       121
    non_face       0.00      0.00      0.00         1

    accuracy                           0.99       122
   macro avg       0.50      0.50      0.50       122
weighted avg       0.98      0.99      0.99       122
```

- **Number of misclassified images: 1**
- **Number of correctly classified images: 121**



b. Alternative Train/Test Split (7 train, 3 test)

New split accuracy:

- **PCA Accuracy (7/3 split): 0.9833333333333333**
- **LDA Accuracy (7/3 split): 1.0**

Comparison with 50/50 split:

- **PCA Accuracy: 0.955**
- **LDA Accuracy: 0.95**

8. Conclusion

This project successfully demonstrated the application of dimensionality reduction and classification techniques for face recognition using the ORL dataset. By transforming high-dimensional image data into lower-dimensional subspaces using PCA and LDA, we were able to efficiently and effectively classify faces with relatively high accuracy. Key observations include:

- **PCA showed strong performance with accuracy increasing as more variance was retained, though it is unsupervised and may not capture class separability as effectively as LDA.**
- **LDA, being supervised, offered improved class discrimination, particularly when using 39 dominant eigenvectors, and often outperformed PCA in classification accuracy.**
- **The K-NN classifier was sensitive to the choice of k, with smaller values like k = 1 or k = 3 generally performing better.**
- **Changing the training/test split affected the model performance, highlighting the importance of balanced and representative datasets.**

Overall, the project provided valuable hands-on experience in image preprocessing, dimensionality reduction, and classification. It also emphasized the trade-offs between model complexity, dimensionality, and classification performance. The results indicate that classical machine learning techniques, when applied thoughtfully, can achieve solid performance in facial recognition tasks.

Source code:
https://github.com/Dalioy/Face-Recognition-Classification-Using-PCA-and-LDA.git