Presentation script

Checkbox	☐

1. **Introduction to the topic**

    a. **Problem definition**

    The human face is considered the prime factor in recognizing a person’s identity in our society. Humans can quickly identify individuals through their visual. The face recognition system was built based on the fact that each person’s face is unique.

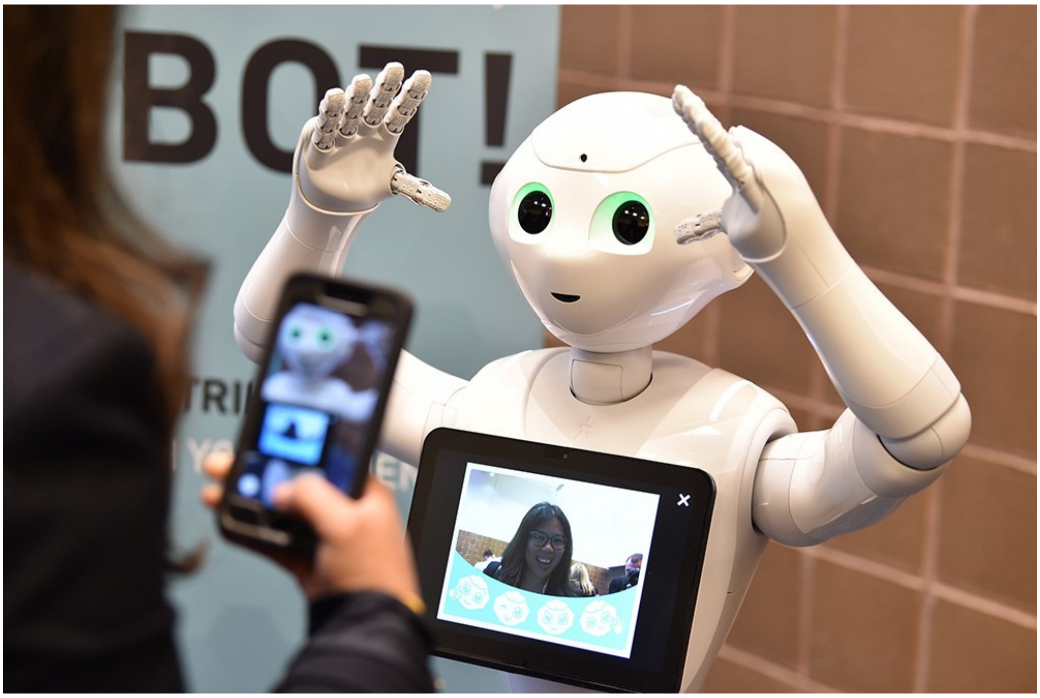
    This technology serves two main purposes: verification and identification. In verification, a face image is matched with a claimed identity to determine its accuracy. For example, if someone says they're John Smith, the system checks if their face matches John Smith's. In identification, a face image is used to determine the person's identity by comparing face with the database of images of known individuals.

    Face recognition finds applications in various fields such as security, surveillance, investigation, identity verification, and more. Its versatility makes it invaluable in enhancing safety, efficiency, and accuracy across different sectors.

    b. **Motivation**

    - Enhance Customer Service by Identifying Customers
        - identify the customer segment based on purchasing/ visiting frequency
        - analyzing customers past purchase history
        - knowing demographic information of customer

      In sectors such as retail and hospitality, we can use face recognition technology at entry points/check-in counters to instantly identify the customer. By identifying, we can automatically extract meaningful information such as how frequently the customer has used our service, their past purchase history, and the demographic information. This enables staff to provide personalized recommendations, enhancing their shopping experience and increasing the likelihood of purchasing.



c. **Problem setting**

- Task 1, 2: **ORL dataset**

- Description
    - All the images were taken against a **dark homogeneous background with the subjects in an upright, frontal position** (with a little side movement)

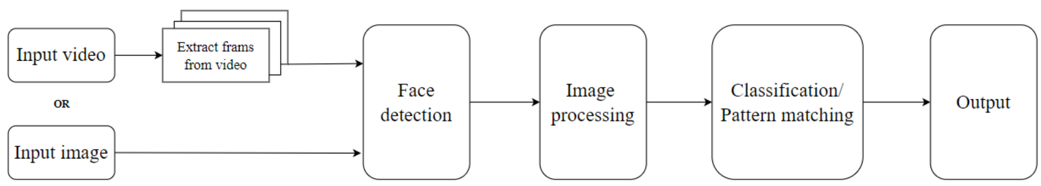
    - The images were **taken at different times, varying the lighting, facial expressions** (open/closed eyes, smiling / not smiling), **and facial details** (glasses / no glasses)
    - **contains 400 images from 40 distinct subjects**
    - **size: 92 × 112**
- Task 3: LFW (Labeled Faces in the Wild) dataset
  // chỗ này nhiều ảnh quá lấy mấy hàng là đc

  - Description
    - desired to study the unconstrained problem of face recognition, such as **variation in posture, facial quality, hairstyles, focus**, and other parameters

    - We use the preprocessed excerpt of the "Labeled Faces in the Wild" from sklearn and **choose images from people that have at least 70 images with size 62×47**
    - **contains 1288 images from 7 identities**
- **Goal: accurately and reliably recognize and verify the identity of individuals based on their facial features under varying conditions**

d. **Demo**

Pipeline

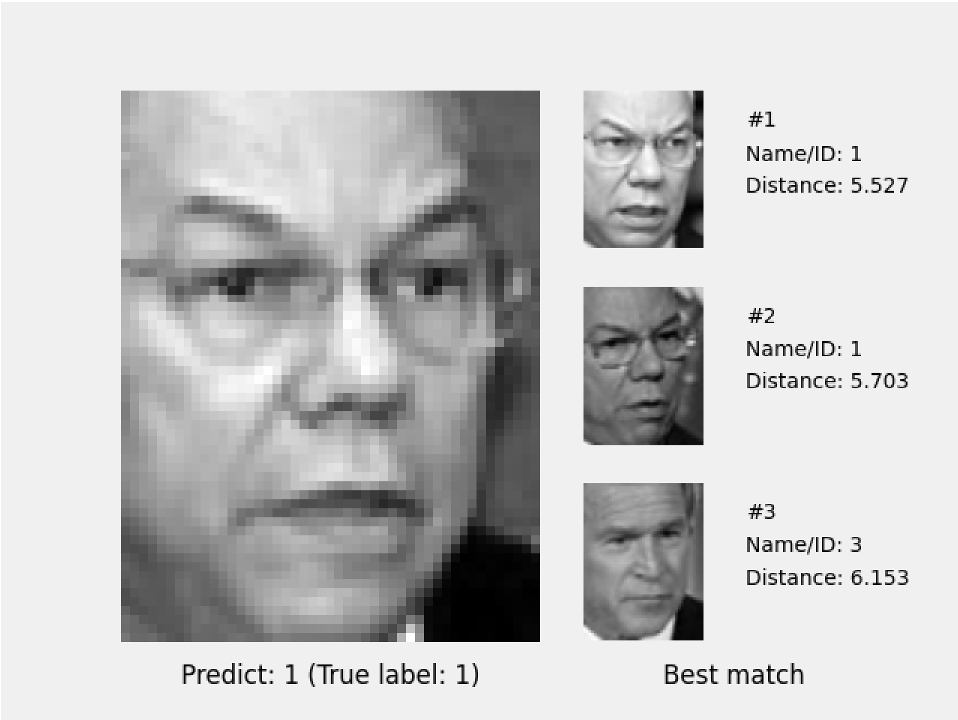


- Face detection: Identify and locate faces within an image or a video stream
- Image processing: Enhance or preprocess the facial images to improve the accuracy of subsequent steps
- Classification/ Pattern matching: Identify the individual in the detected face(s) by comparing facial features to known faces
- Output: Provide the identified name of the person in the image or the identity that best matches the detected face

Real-time

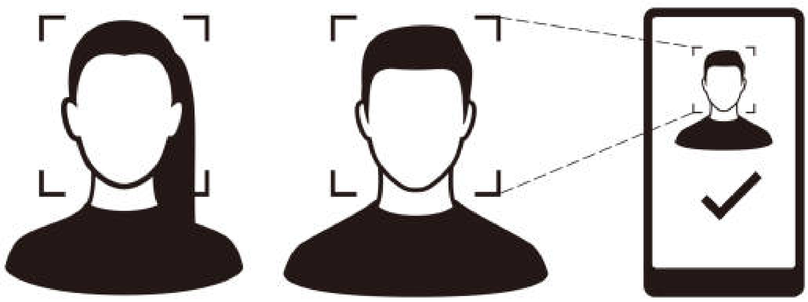
// video để gửi qua zalo

Prediction & Best match



2. Literature review(phần này đang sửa lại nhá)
    - 2 step of recognizing person’s face

    The field of face recognition research has been experiencing rapid advancements. **According to M.K.Rusia (2022), the process of recognizing a person's face can be broken down into two steps: first, detecting the location of the face, and second, classifying the detected region.**



Over the last decade, numerous approaches have been introduced for face recognition, which can be categorized into four types:
appearance-based, feature-based, model-based, and hybrid approaches. Today, **deep learning technology** is becoming increasingly popular due to its ability to recognize patterns and handle variations.

Several approaches have been experiment with **ORL database** and achieve significant result. The **eigenfaces method** introduced by **M. Turk and A. Pentland (1991)** produced **89.5% recognition rate. Lawrence et al (1997)** also reported a **96.2% recognition rate** using **hybrid neural network solution**. **LFW dataset** is also a widely used benchmark for face recognition algorithms, **D. Yi et al (2014)** achieve **accuracy of 96.33% after tuning with PCA**. **Taigman et al. (2014)** presented **DeepFace**, a deep learning model that achieved an impressive **accuracy of 97.35%** on the LFW dataset

Chellapa et al (1995) presented a survey on several statistic, machine learning and deep learning methods **for face recognition**. **Currently, one of the methods that yields promising results on frontal face recognition is the principal component analysis (PCA).** PCA can:

+

Reduce the number of variables in a dataset while retaining the most important information

+

Improve the interpretability of the data and make it easier to understand

+

Simplify high-dimensional data visualization, and **reveal patterns** and relationships.

Therefore, we decided to use PCA with some kinds of distances and classification methods to build a face recognition system

The main objective of this research is to improve the accuracy and precision of face recognition under varying facial expression, illumination, and pose conditions in ORL and LFW dataset by
modular PCA method with **advanced image processing techniques** such as **Contrast Adjustment, Bilateral filter, and Gaussian Blur.**

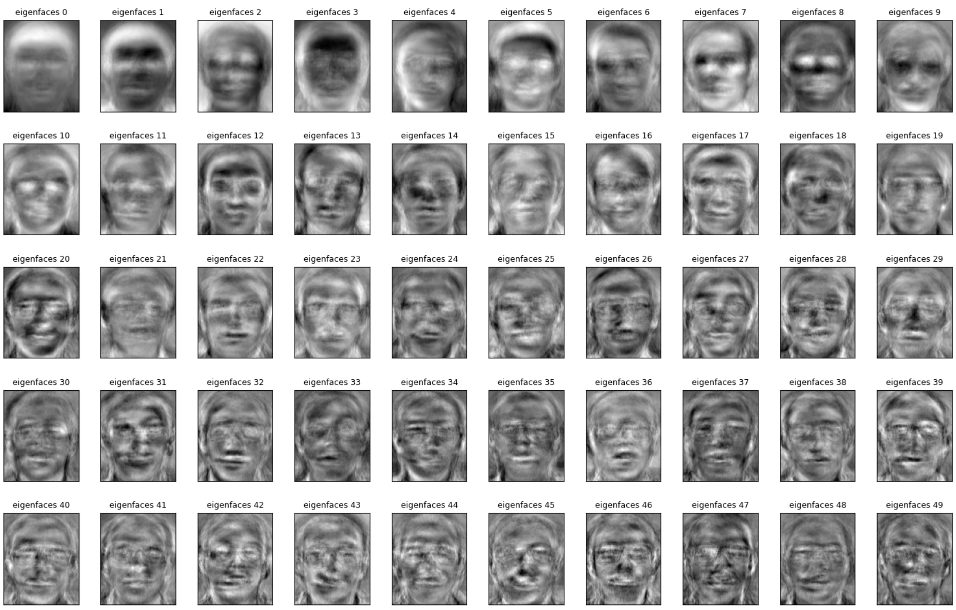
As you can see in the line chart as we show here, **the best combination is Incremental PCA combined with Minkowski distance.**

3. Solution: detailed description

▼ Task 1

Perform PCA to reduce dimension to 50 components

Eigenface:



- Each face can be represented as a linear combination of the eigenfaces. Each face can also be estimated using the 'best' eigenfaces, which have the largest eigenvalues and represent the largest variations in the face image database. The images of each person are stored in individual folders. If a person is selected as the input to run the program, then the folder belonging to that person is selected.

—

For face recognition, in the first task, we have tried using **6 different types of PCAs, combined with 14 different types of distance measures**. From each pair of combinations, we obtain a list of results. From there, we compare them with each other, choosing the combinations that give the best results for each type of PCA. (Show chart...)

As u can see, modified SSE is suitable for most types of PCA algorithms. But the best result is incremental PCA combined with Minkowski distance. It is understandable.

- **Incremental PCA** is a variation of traditional PCA *designed to handle large datasets that may not fit into memory at once*. Incremental PCA processes the data in smaller batches or chunks, making it more memory-efficient and scalable to large datasets.
- **Minkowski** is a (versatile) distance metric that includes both Euclidean and Manhattan distances as special cases. It can be tuned using the parameter p to adjust sensitivity to different dimensions.

▼ Task 2.

2. Evaluation Different Classification Model

When trying to find the most efficient model for face recognition on this dataset, we have experimented 6 familiar ML Classification Models include : Logistic Regression, SVM, Naives Bayes, KNN, Decision Tree, Random Forest. Most of them provide a good performance on small dataset like ORL.

Specifically, 2 models Logistic Regression and SV after tunning hyperparameter provide a 100% accuracy recognition.

Model	Accuracy	Recall	F1-Score	Precision	ROC AUC
1. SVM (linear) + PCA	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**
2. Decision Trees + KernelPCA	0.7375	0.761261	0.718112	0.755405	0.887639
3. K-Nearest Neighbor + PCA	0.9875	0.993	0.992	0.993	0.996
4. Gaussian Naive Bayes + Gaussian Naives Bayes	0.975	0.9583	0.9506	0.9528	0.979
5. Logistic Regression+ PCA	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**
6. Random Forest + PCA	0.9875	0.9722	0.9667	0.963	0.9859

▼ **Logistic Regression and Support Vector Machine provide the best efficient for a small dataset like this**. This might due to:

- For LR : Linear separability: While not always the case, in the ORL dataset, the facial features of different individuals may exhibit some degree of linear separability. This means that a linear decision boundary, as learned by LR, can effectively separate different classes (individuals) in some cases. And L2 Regularization
- For SVM: Kernel trick (SVMs): While the base SVM assumes linear separability, it can be enhanced by using the kernel trick.

This allows SVM to learn non-linear decision boundaries in the high-dimensional feature space, potentially improving performance even if the data exhibits some non-linearity.

- Decision Tree : provide poor performance since its inability to handle with high dimensional data and non-linear relationship,
- Gaussian Naive Bayes: provide poor performance since it assumes all features are linearly independent → which hard to exist in face recognition

Wrong Prediction Image in Test set:

Wrong Prediction Face:

- **Wrong prediction tend to occur at person contains characters like wearing glasses, having beard or mustache, different head position, facial expression, or lightning condition.**
- These wrong predicition images often observed in model such as Decision Tree, Random Forest and Naive Bayes since they inability to handle with high dimensional data, or non-linear relationship and more likely to overfitting

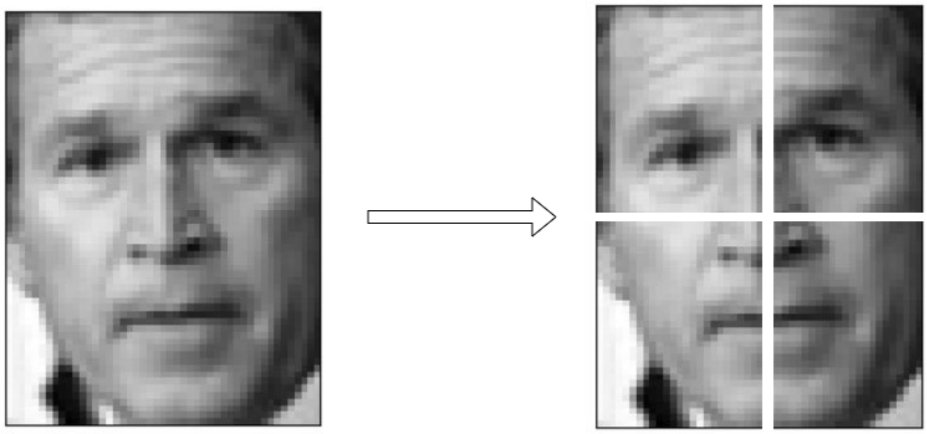
→ **We gonna use Logistic and SVM for afterward deployment**

▼ Evaluation in Different Types of PCA

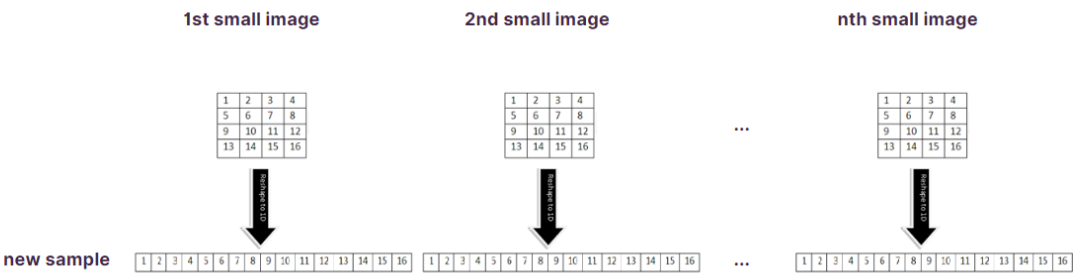
....

▼ Task 3

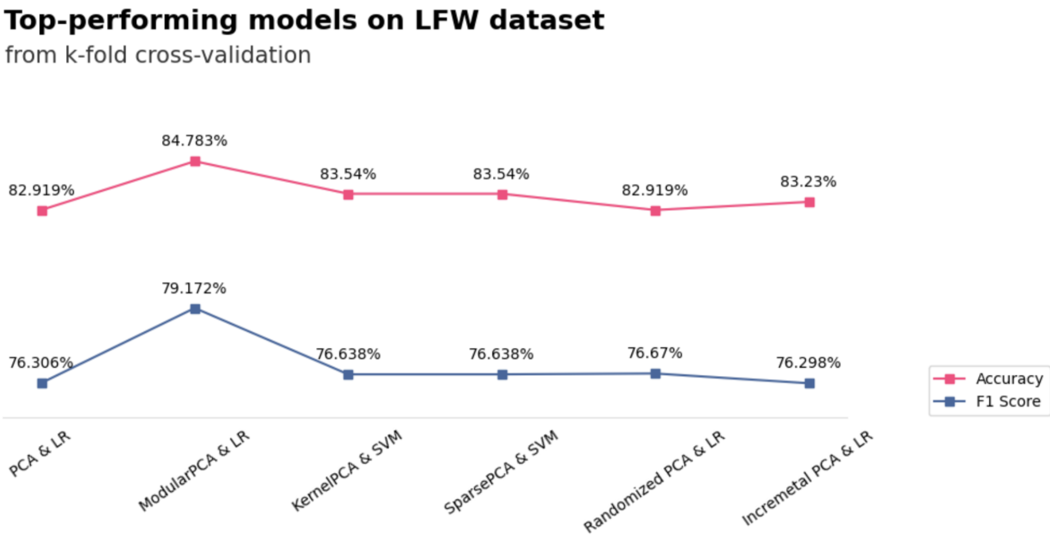
Modular PCA is an extension of the conventional PCA method
+ In the modular PCA method, the
face images are divided into smaller images that are equal in size. Each small image is represented as a numpy matrix.



+ **Reshape each image as an array and concatenate the array of these small images to a sample**



+ Then we **apply conventional PCA to get the result**



As we can see here, Modular PCA outperforms other methods by at least 1% in score. Therefore we decided to use Modular PCA as the main method to reduce dimension.

Also, we chose Weighted SSE and Logistic Regression to predict labels due to their highest results.

Next, we will focus on image enhancement to improve scores.

Image enhancement is the process of improving the quality and appearance of an image. In this research, we focus on 3 techniques

1. **Contrast Adjustment**

   → *involves modifying the difference in intensity between the light and dark regions of an image*

   The basic formulation for contrast adjustment can be expressed as:
   Adjusted Pixel Value=
   $\alpha \times$Original Pixel Value$+\beta$

   Here:

   - α is the contrast control parameter.
   - β is the brightness control parameter.
   - Original Pixel ValueOriginal Pixel Value is the intensity value of a pixel in the original image.
   - Adjusted Pixel ValueAdjusted Pixel Value is the resulting intensity value after contrast adjustment

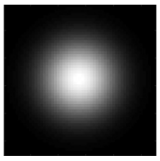
Low contrast Original High contrast

The main idea: The matrix of images is multiplied by alpha and then the pixel values are clipped to be between 0 and 255.

2. **Gaussian Blur**

→ is used for smoothening images and reducing noise

$$GB[I]_p = \sum_{q \epsilon S} \underbrace{G_\sigma(||p-q||)}_{\text{Normalized Gaussian Function}} I_q$$



Here, $GB[I]p$ is the result at pixel *p*, and the RHS is essentially a sum over all pixels *q* weighted by the Gaussian function. I_q is the intensity at pixel *q, it* refers to the grayscale value or color value of a pixel in an image

GaussianBlur(radius=5)

***Parameters**: **radius** – blur radius. Changing the value of the radius the different intensities of the GaussianBlur image were obtained The higher the radius the blurrier the image is*

Original

Gaussian Blur

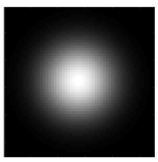
To apply the Gaussian blur to an image, the 2D Gaussian kernel is convolved with the image matrix. The convolution operation involves sliding the kernel over the image and computing the weighted sum of the pixel values in the neighborhood defined by the kernel at each position.

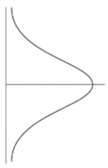
$$G(x,y) = \frac{1}{2\pi\sigma^2} \cdot e^{-\frac{x^2+y^2}{2\sigma^2}}$$

3. **Bilateral filter**

→ is used for smoothening images and reducing noise, while preserving edges

$$BF[I]_p = \underbrace{\frac{1}{W_p}}_{\text{Normalization Factor}} \sum_{q\epsilon S} \underbrace{G_{\sigma_s}(||p-q||)}_{\text{Space Weight}}\underbrace{G_{\sigma_r}(|I_p - I_q|)}_{\text{Range Weight}} I_q$$





OpenCV has a function called **bilateralFilter()** with the following arguments:

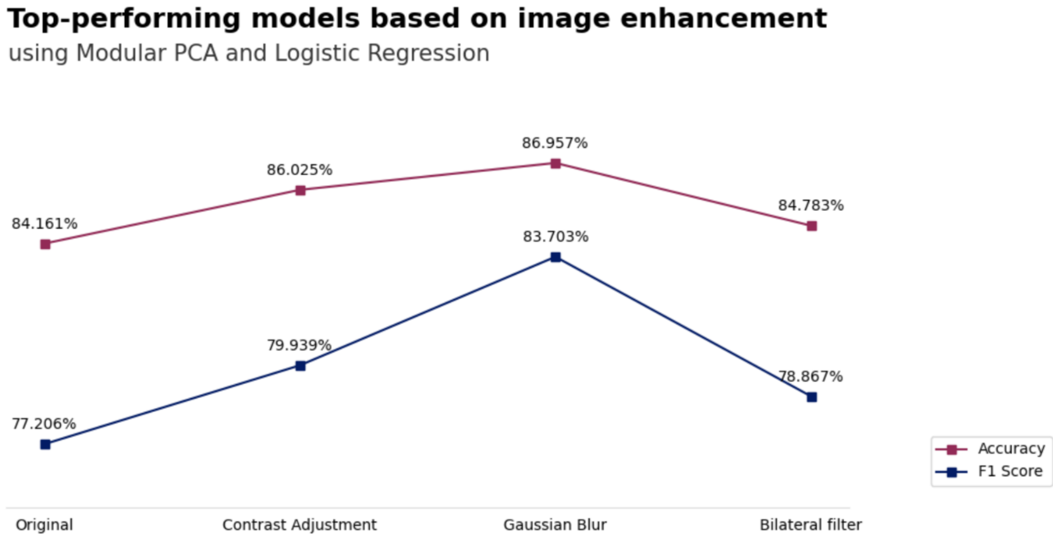
1. **d:** Diameter of each pixel neighborhood.
2. **sigmaColor:** Value of in the color space. The greater the value, the more colors farther from each other will start to get mixed.
3. **sigmaSpace:** Value of in the coordinate space. The greater its value, the more further pixels will mix, given that their colors lie within the sigmaColor range.

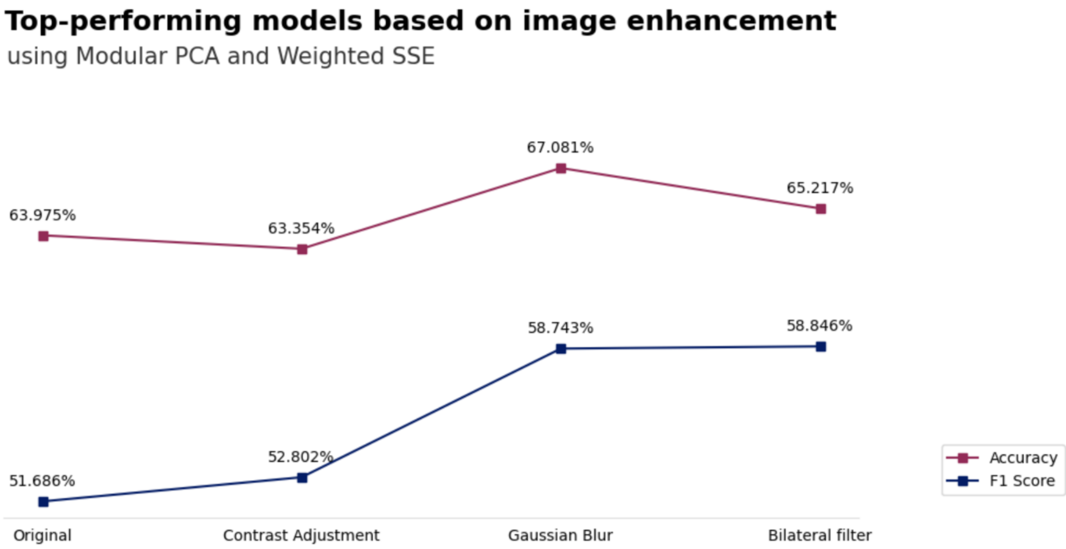
Original

Bilateral Filter

Compare to GaussianBlur, the Bilateral filter has the normalization factor and the range weight as new terms

The bilateral filter considers both the spatial proximity and intensity similarity between pixels when determining the smoothing effect. It ensures that only those pixels with intensity values similar to that of the central pixel are considered for blurring, while sharp intensity changes are maintained





In this case, both image enhancement techniques can remove noise, filter to get the important information, and therefore improve score and the Gaussian blur yields the best outcome.
We can conclude that image enhancement techniques can improve

4. Result: Metrics of evaluation

   - Accuracy is a measure of the overall correctness of the model. It is calculated as the ratio of correctly predicted instances to the total instances. Formula:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$

- F1 score is the harmonic mean of precision and recall. It provides a balance between precision and recall, especially when there is an uneven class distribution. Formula:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

- ATT dataset
  - Best score: Accuracy 1.0, F1 1.0
- LFW dataset
  - Best score: Accuracy 0.86957, F1 0.83703

5. Conclusion

   We have presented the face recognition experiments using different kinds of PCA, distances, classification, and image enhancement processes. Our results indicated that Logistic Regression and Support Vector Machines were more effective than other methods when analyzing the ORL dataset. Moreover, Modular PCA outperformed other techniques when analyzing the LFW dataset. Lastly, we found that image enhancement processes resulted in improved scores compared to the original dataset. The future work will focus on detecting unknown faces and working with 3D images.