Journal of Economics and Administrative Sciences
2024; 30(143), pp. 405-419

P-ISSN 2518-5764
E-ISSN 2227-703X

Journal of Economics and Administrative Sciences (JEAS)

Available online at http://jeasiq.uobaghdad.edu.iq
DOI:https://doi.org/10.33095/2k5d7267

Face Image Recognition by using some Dimension Reduction Algorithms and Logistic Regression

Sura Sabah Keiteb*
Department of statistic
College of Administration and Economics,
University of Baghdad, Iraq
sora.sabbah1101a@coadec.uobaghdad.edu.iq
https://orcid.org/0009-0007-1284-7662
***Corresponding author**

Entsar Arebe Fadam
Department of statistic
College of Administration and Economics,
University of Baghdad, Iraq
entsar_arebe@coadec.uobaghdad.edu.iq
https://orcid.org/0009-0000-0316-6995

Received: 7/9/2023 Accepted:7/2/2024 Published Online First: 1 /10/ 2024

Abstract:

Face recognition is a common technique for artificial intelligence and image processing in which machine-learning algorithms automatically recognize a person's face through a huge image collection or even from a live video since there is difficulty in detecting data when using different shots for the same person in a facial recognition system, it takes time to determine person's membership. Furthermore, when working with high-dimensional data, where the 2D-pixel values of facial images, can be computationally challenging, the face recognition problem gets harder, so the current research compares the dimensionality reduction techniques like the principal component analysis algorithm PCA, an unsupervised learning algorithm that reduces dimensions by moving them from a high-dimensional area to a lower area without losing important information, and the two-dimensional principal component analysis algorithm 2DPCA to determine which technique is most effective for the process of detecting facial images, and this method will be compared with the logistic regression which is a statistical model for face recognition that has been used for classification tasks. Two different sets of data were used, the first group represented the ORL database, which was made up of 40 individuals, while the second group represented the real data, which was made up of 100 individuals from AL-Mamoon College, the results showed that the logistic regression model achieved the lowest MSE for ORL dataset and real data by achieving (0.0024414, 0.091936) respectively, followed by 2DPCA and PCA and all three techniques reached the highest accuracy rate of 100% for facial image recognition on real data.

Paper type: Research paper

Keywords: Logistic Regression, Principal Component Analysis, Two Dimension Principal Component Analysis, Face Recognition, Pattern Recognition.

1. Introduction:

Face recognition is a computer vision technique uses a digital image or video feed to identify or confirm a person's identification. Since smartphones, security cameras, and facial recognition software have grown in popularity, this procedure has gained more traction, it takes a multifaceted strategy integrating knowledge of computer vision and machine learning to address the problems with face recognition.

FR a crucial and challenging topic at the same time due to the impact of facial images by various factors in various circumstances. such as dealing with variations in lighting conditions, distinguishing faces in cluttered backgrounds, identifying faces of different ages and ethnicities, recognizing partially visible faces, dealing with disguised faces, handling expression variation, rotation, and others.

The ability to strike a balance between these conflicting challenges and guarantee dependable and accurate face recognition is ultimately what determines whether face recognition systems succeed. Continued progress in the creation of machine learning algorithms is necessary to reach this degree of success. Because of that importance, a variety of statistical approaches have been developed to resolve these types of problems, including principal component analysis (PCA), which is a popular statistical technique for dimension reduction technique used in machine learning and data analysis. It transforms the high-dimensional feature space into a lower-dimensional feature space, capturing the most significant features of the data.

Recently, Yang suggested two-dimensional principal component analysis (2DPCA) as a different facial recognition approach is an extension of PCA to handle two-dimensional data, such as images. It finds the principal components of a two-dimensional dataset and can be used to improve face recognition accuracy.

The 2D image matrix is handled instead of converted into a 1D vector, in contrast to PCA (Frangi and Yang, 2004), therefore compared to PCA, we may achieve higher computation accuracy and a considerably lower covariance matrix. Numerous common techniques utilized as classifiers have recently been suggested, including the logistic regression model, which is a machine-learning classification algorithm. It simulates the relationship between the dependent variable (binary) and one or more independent variables (numerical). And often employed in face recognition tasks to categorize face images into classes or generate class scores.Figure 1 shows the process of face recognition.

To finish this research, we divided it into four chapters: the first chapter contained the introduction, the second chapter represented the material and methods, the third chapter represented the discussion of results and the fourth chapter contained the research's conclusions.

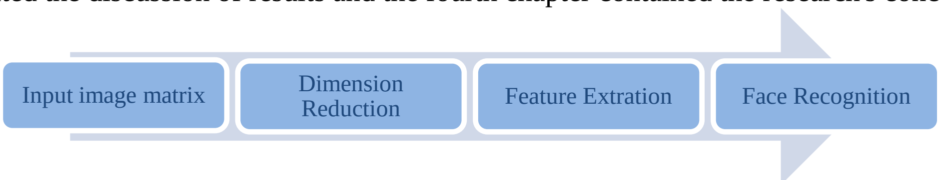

Figure 1: Process of the face recognition system

1.1 Literature Review:

In this section, dimension reduction algorithms and logistic regression were being used for face recognition tasks as follows:

Soni and Sahu (2013) used the 2DPCA technique, a standard matrix-based feature extraction technique in which the face image arrays were utilized in both directions, and eigenvalue-based projection vectors were compared using measures of Euclidean distance, city block distance, Mahalanobis distance, and covariance similarity. The proposed method merges the two processes of classifier-based recognition and extraction of features through 2DPCA. For the experiment, the ORL face database was employed, this database contains images with various positions, illumination, mustaches, beards, and eyeglasses, the best results were obtained with 2DPCA and city block distance.

Zhou et al (2014) proposed a novel face recognition method that is based on PCA and logistic regression. Their method used PCA to identify features and minimize the dimensions of process data. They then introduced a novel classified approach and employed logistic regression as a classification method for facial recognition. The results of the experiment on two separate face databases were provided to show the efficacy of their proposed strategy.

Abdulmunem and Ibrahim (2016) introduced a powerful face recognition technique based on wavelet decomposition to extract the most significant and distracting features for the face and the eigenface approach to classify faces based on the shortest distance between feature vectors. The strategy was tested using the faces 94 database, with an accuracy of 100% and a recognition time of 87.5%, an outstanding recognition with minimal computing time was produced.

Vanlalhruaia et al (2017) suggested applying logistic regression and a neural network to recognize faces in binary photos. The technique suggested transforming a color image to grayscale and then denoising it using a low-pass filter. The denoised image was then separated using adaptive thresholding to detect the local intensity fluctuations around the brows, eyelids, nose, and mouth. The binary image was subsequently normalized and reduced to 50%, 30%, 20%, and 10% of the initial measurement using the nearest neighbor interpolation technique, and a face database was developed for every image. The logistic regression method yielded 100% training and testing accuracy, whereas the 3-layer Back-propagation neural network (BPNN) yielded 100% training and 97.5% testing accuracy even when the size was reduced to 20% of the original image dimension.

Ma and Yuan (2019) focused in their paper on dimension reduction research. When extracting features from images with deep learning, the PCA approach was applied to reduce dimension. They started by using deep convolutional neural networks to identify image features. The PCA algorithm was subsequently implemented and used to solve the challenge of processing high-dimensional sparse large data, during image pre-processing, simulation tests were used to validate the feasibility of the PCA approach for dimension reduction of image feature extraction by deep learning. The suggested algorithm's efficiency was demonstrated by comparing its efficacy to that of PCA.

Wang et al (2021) created a new approach to facial image recognition named truncated nuclear norm on low-rank discriminant embedding (TNNL). The TNNL could reduce the adverse consequences of noise while improving feature distinction. They also presented two iterative techniques for extracting the resilient low-dimensional image feature. To prove the usefulness and resilience of TNNL, they ran experiments on two benchmark face image databases for low-dimensional feature extraction. The experimental results showed that TNNL outperformed the previous approaches.

Beiranvand et al (2022) indicated an extensive unsupervised feature selection method called unsupervised feature selection using principal component analysis (UFSPCA), in which they initially utilized PCA to construct unrelated and orthogonal features, then computed the correlations between the original and unrelated features, then analyzed two sets of their initial and orthogonal traits, as well as how comparable they were to a weighted bipartite graph. Finally, using the Hungarian method, they found a match with the largest weight. To demonstrate the optimality and efficiency of the proposed method, they tested the method on five datasets using the KNN classifier by comparing it to seven well-known unsupervised feature selection algorithms. The results of the study revealed that the UFSPCA approach outperformed the other seven methods.

Shanthi et al (2023) illustrated facial recognition techniques used in drones include linear discriminant analysis (LDA), local binary pattern histogram (LBPH), principal component analysis (PCA), elastic bunch graph matching (EBGM), and neural networks. Their research would aid facial recognition technologists in developing an algorithm that is a hybrid based on the requirements of applications in real time.

P-ISSN 2518-5764
E-ISSN 2227-703X

The problem of this research is to identify a person's face from several images and when there are different shots of the same person which is a crucial problem in practical scenarios., especially when working with high-dimensional image data, images may differ significantly when a person's face is captured from different angles or in changing lighting. This makes it challenging for the face recognition system to correctly identify the person. So, the objective of this research is to compare the three methods (PCA, 2DPCA, and Logistic Regression) together in a facial recognition system to identify the appropriate method of facial image recognition to manage it in two steps: first, Applying PCA and 2DPCA, reduces the dimensionality of the input face image data, the lower-dimensional data that results can then be used as input for the face recognition step, and then use the facial recognition system's classifier, logistic regression on the original face image data.

2. Material and Methods:

2.1 Data set for the Research Topic:

In the age of digital technology, facial recognition has become a crucial challenge in pattern recognition (Wang et al, 2021), and a vital part of everyday life for people, as this framework will help us to identify people by passing through the view of the reconnaissance camera, face recognition is generally carried out through a digital image or video clip to study and compare these patterns. Where it supports face recognition numerous application fields, including digital forensics, biometrics, banking and finance, mobile technology, security services, etc., have benefited greatly from the rapid expansion of facial recognition technologies. To identify an individual, facial recognition systems simulate human face understanding and recognition abilities (Pattnaik and Mohapatra, 2023).

Various beneficial image recognition of the face techniques subspace learning-based have been developed during the past few decades, nevertheless, these techniques may efficiently eliminate the unnecessary data in the dimensions image space for input, by projecting the image of the face into a low-dimensional subspace, it may also extract an accurate low dimensional image feature. In some sense, all subspace learning-based approaches are dimensionality reduction techniques, where unnecessary data can be lowered without no sacrificing important image data (Wang et al, 2021). Due to the high dimensions of this image, we need to use algorithms to reduce the dimensions in facial image recognition. These algorithms are considered unsupervised machine learning techniques and have been successfully applied to face recognition such that PCA and 2DPCA (Shanthi et al, 2023), compared to logistic regression, the most common and widely applied supervised learning algorithm in machine learning for face recognition, the data on the subject of the research represent the individual's images from two experiments, and we can illustrate as follows:

1. The first group: This group represents the database of ORL (Olivetti and Oracle Research Laboratory), a database used for the experiment and face recognition, this data includes 40 individuals as shown in figure 2, and 10 different images were taken for each individual with a size of 112×92 and 10304 pixels as shown in figure 3.

Figure 2: Individuals from the ORL dataset

Figure 3: Ten different images of one individual from the ORL dataset

2. The second group: This group represents the real data that was collected from 100 students at AL-Mamoon College (Department of Business Administration) as shown in figure 4, during 4 months using a digital camera to take color images from various angles. Additionally, 10 different images were taken for each student, as shown in figure 5.

Figure 4: Individuals from real data

Figure 5: Ten different images of one individual from real data

The color images for each individual are converted into gray images to deal with the 2D matrix and all of the images are resized to be the same size, with dimensions ranging from 112x92 and 10304 pixels, in this way, we can deal with a two-dimensional digital matrix m×n. As there are numbers (pixel values) inside each data matrix, these numbers which range from 0 to 255, where 0 represents the color black and 255 the color white, might indicate the brightness of the light or the precision for the clarity of the image.

2.2 Dimension Reduction:

In data mining and machine learning, dimension reduction approaches are strategies used to minimize the number of dimensions in a dataset while maintaining the most pertinent information. These techniques are especially helpful when working with high-dimensional datasets, which can be challenging to handle and evaluate by extracting the most relevant features from the data (Fırat et al, 2022).

2.2.1 Feature extraction:

This is the first step in the process of dimensionality reduction, the input data is converted into a feature vector, which is an expanded version of a set of features. This step fixes the issue of the high dimensionality of the input dataset. This means that only those relevant features are extracted instead of the entire dataset. Once the features have been extracted, select the ones that have the best chances of providing correct results (Ebied et al, 2012).

2.2.2 Principal Component Analysis :

The principal component analysis algorithm is a popular statistical tool for current data analysis and an unsupervised learning algorithm in machine learning. By using measurements such as the arithmetic mean, standard deviation, covariance matrix, eigenvalues, and eigenvectors (Prusty et al, 2017), PCA linearly reduces a collection of information into lower dimensions known as principal components PCs by geometric projection. (AL-Rawi and Isa, 2019), and to achieve the required information for the data set, the principal components (PCs) are orthogonal to each other (Shanthi et al, 2023). The components with larger variance can provide a lot of information, whereas the components with lower variance can only provide little data (Karthick et al, 2021).

The importance of This application is attributed to its employment in numerous significant research fields, including pattern recognition, dimensional reduction, feature extraction, image compression, and other areas. The basic idea behind the face recognition system using PC is converting each training image into a vector, a predetermined set of training images is used to generate the covariance matrix., and the eigenface is the eigenvector acquired from PCA (Abdulmunem and Ibrahim, 2016). Its operating model will be as follows (Ma and Yuan, 2019):

If we have a set of data, which is denoted by X, with high dimensions X=$\{x_1, x_2, x_3, \ldots, x_n\}$, we can through PCA convert this data from its high dimensions to lower dimensions where the X_i represents the sample i, observation n represents the total number of observations or samples (Tharwat, 2016), and all samples have the same dimension ($x_i \in R^m$).

To calculate the eigenvalues and vectors (Beiranvand et al, 2022), we will first calculate the arithmetic mean (μ) in the data matrix for each variable according to equation (1):

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} \quad (1)$$

Then we calculate mean-centering data, which is denoted by D by subtracting the average for every sample as follows:

$$D = \{d_1, d_2, d_3, \ldots, d_N\} = \{x_1 - \mu, x_2 - \mu, \ldots, x_N - \mu\}$$

Following that, the covariance matrix is computed as equation (2):

$$\Sigma = DD^T \quad (2)$$

The covariance matrix is a symmetric matrix with a dimension M×N and a (p.s.d) matrix, meaning that all eigenvalues of X are greater than or equal to zero.

The covariance matrix's eigenvalues and eigenvectors are then calculated according to equation (3):

$$V = \mathrm{eig}(\Sigma) \,.\, \lambda \quad (3)$$

Where V and λ denote eigenvectors and eigenvalues, respectively, with a dimension M×N, and according to the properties of the covariance matrix, the eigenvectors that represent the principal component corresponding to the eigenvalues are orthogonal. The eigenvector with the highest eigenvalue can be assumed as the first principal component PC and has the greatest variation in comparison to the other principal components. As a result, a new matrix is created, and let's assume Y where the initial data set X with a dimension M×N is converted to a new matrix Y with Dimension M×K.

The following steps can be used to explain the steps of the PCA algorithm:

Step 1: The data or feature matrix X with a dimension (M × N) is generated in the PCA algorithm. which is considered as a first step, where in the matrix one column represents the sample while the rows represent the dimension.

Step 2: Convert the image matrix of X with 2D (112×92) where 112 represents the rows and 92 represents the columns to image vector 1D.

Step 3: Calculate the mean ($\mu_{M\times1}$) of all samples of the X matrix is calculated as follows:

$$\mu = \frac{1}{N} \sum_{i=1}^{N} x_i$$

Step 4: Subtract the mean of all samples from each image vector as follows:

$$D = \sum_{i=1}^{N} x_i - \mu$$

Step 5: Calculating the covariance matrix ∑ as follows:

$$\Sigma = \frac{1}{N-1} DD^T$$

Step 6: Compute the covariance matrix's eigenvalues (λ) and eigenvectors (V) as follows:

$$V = \mathrm{eig}\,(\Sigma) \,.\, \lambda$$

Step 7: The eigenvectors that correspond to the eigenvalues are arranged.

Step 8: Create an eigenface (S) by multiplying the mean subtracted with the eigenvectors as follows:

$$S = D.\,V$$

Step 9: Keep only the k Eigen faces corresponding to the k eigenvectors with the largest eigenvalues.
Step 10: After completing the above procedures, the image to be identified is compared with the database's image collection, and the matching result is obtained by obtaining the absolute value.
Step 11: The image belongs to the group of images where the absolute value is zero.

2.2.3 Two-Dimensional Principal Component Analysis :

The two-dimensional principal component algorithm was developed by Yang and it is widely used in many applications, including dimensional reduction, facial representation, and recognition. This technique can be considered an expansion of the traditional PCA technique, and its outcomes are more precise and effective than PCA (Bengherabi et al, 2008).

The main idea behind this method is that instead of using 1D vectors for 2D images, it uses two-dimensional matrices, generates the two-dimensional covariance matrix from the image matrices, and decreases the image covariance matrix, this approach does not require the image matrix to be converted into a vector. (Zhang and Zhou, 2005), compared to classic PCA, 2DPCA can more properly evaluate the matrix, calculate it more quickly, and find out the eigenvectors that correspond to the eigenvalues (Bengherabi et al, 2008). The principle of its work is as follows:

The basic idea of the 2DPCA technique is to project the image A onto the vector X, where A is a random image matrix of dimension m and X is the projection vector of dimension n (You and Cai, 2009). Through this projection, we can obtain the projected Vector Y by the subsequent linear transformation as equation (4):

$$Y=AX \qquad (4)$$

As a result, a projection vector Y with a size m is created, and this vector is known as the projected feature vector of image A. The total dispersion of projected samples is required to produce a decent projection vector X, and the performance of the projection vector X may be gauged using the following criterion, which can be calculated using the equation below:

$$\begin{aligned} J(X) &= \text{trace }\{E[(Y-EY)(Y-EY)^T]\} \\ &= \text{trace }\{E[(AX-E(AX))(AX-E(AX))^T]\} \\ &= \text{trace }\{X^T E[(A-EA)^T(A-EA)]X\} \qquad (5) \end{aligned}$$

We can determine the projected vector X by solving the eigenvector of the covariance matrix (scatter) of the image, which is denoted by G and can be calculated according to equation (6):

$$G=E\,[(A-E(A))^T\ (A-E(A))] \qquad (6)$$

It is a specific non-negative matrix with a dimension m, and assuming that we have training samples M, the covariance matrix G can be expressed as equation (7):

$$G=\frac{1}{M}\sum_{k=1}^{M}(A_k-\bar{A})^T(A_k-\bar{A}) \qquad (7)$$

A_k represents the training sample matrices with dimension m and A_k (k=1, 2…, M) It represents the average image of samples for M training, which can be calculated according to equation (8):

$$\bar{A}=\frac{1}{M}\sum_{k=1}^{M}A_K \qquad (8)$$

The optimal solution for the projection vector is X_{opt} formed by the orthogonal eigenvectors $x_1,x_2,\ldots,x_d$ of the covariance matrix G corresponding to the largest eigenvalues
The following steps can be used to explain the steps of the 2DPCA algorithm:
Step 1: Determine the matrix's data for image A, which has m×n dimensions.
Step 2: Compute the average image of samples for M training images as illustrated in the following:

$$\bar{A}=\frac{1}{M}\sum_{k=1}^{M}A_K$$

Step 3: Subtract the average of all samples from each image matrix.

P-ISSN 2518-5764
E-ISSN 2227-703X

Step 4: Compute the covariance matrix G which is a square matrix n×n as illustrated in the following:

$G=\frac{1}{M}\sum_{k=1}^{M}(A_k-\bar{A})^T(A_k-\bar{A})$

Step 5: Decomposition of the matrix and calculation of the eigenvalues and eigenvectors.

Step 6: Ordering eigenvectors in descending order, a diagonal matrix corresponding to eigenvalues.

Step 7: The vector with the highest eigenvalues is selected and its value is multiplied by the original matrix A and given new coordinates as follows:

Y=AX

Step 8: Through the steps mentioned previously, the database's image collection is compared with the image that needs to be identified, and the matching result is obtained by finding the absolute value

Step 9: The image falls under a class of images whose absolute value is zero.

2.3 Logistic Regression Model :

The logistic regression model is based on the fundamental supposition that the dependent variable (Y) or response variable follows the Bernoulli distribution (Khalaf and Mohammed, 2023), where a binary variable has a value of one (1) and the probability (p) of the response occurring, and has a value of zero (0) and the probability of q= (1-P) when the response does not occur (Saeed, 2015). The response function (logistics function) can be expressed as follows:

$$f(z)=E\,((Y|Z)=\frac{e^z}{1+e^z} \qquad (9)$$

The classification process is dependent on the outputs of this function, which is similar to the theory of probability in that values can be either zero or one. If the function's value is equal to one, the observation belongs to the first group and if it is equal to zero, the observation belongs to the second group and here represents the variable Z independent variables, and this variable measures the sum of the contributions of all independent variables used in this model, which is known as (logit), so the input of the response function is from (-∞ to +∞) while the value of the output is between zero and one (Khairunnahar et al, 2019).

The variable Z in the case of a single independent variable can be defined as equation (10):

$$Z=B_0+B_1X_1 \qquad (10)$$

In the case of more than one independent variable, that is, in the case of multiple logistic regression, it can be written as equation (11):

$$Z=B_0+B_1X_1+B_2X_2+B_3X_3+\ldots\ldots+B_kX_k \qquad (11)$$

B_0, B_1, ..., B_k represent the parameters of the model.

We can by transforming the logit transformation, which is used to convert formula (9) to linear form, which can be illustrated as equation (12):

$$g(z)=Ln\left[\frac{f(z)}{1-f(z)}\right]= B_0+B_1X_1+B_2X_2+B_3X_3+\ldots\ldots+B_kX_k \qquad (12)$$

There are several kinds of methods for estimating the parameters of the logistic regression model, but in this study, we will focus on the most popular technique, known as the maximum likelihood method (Al-Nasrawi, 2017):

By using logistic regression, it is possible to define the likelihood function as the probability of observing a specific pattern of occurrence of the desired attribute (Y=1) or non-occurrence (Y=0) in a sample. When Y=1 for case (i), the probability of getting the observed data is given by the term P(Xi), whereas when Y=0 for case (i), the likelihood of getting the observed data is given by the term 1-P(Xi) as (Al-Jubouri, 2018):

$P(Y=1|X)=f(X) \qquad$ if Y=1

$P(Y=0|X)=1-f(X) \qquad$ if Y=0

When yi follows a binomial distribution, the greatest possibility can be expressed as equation (13):

$$L(B, Y) = \prod_{i=1}^{N} \frac{n!}{y!(n!-y!)} p(x_i)^{y_i} (1 - p(x_i))^{ni-yi} \quad (13)$$

The above equation can be simplified:

$$L(B, Y) = \prod_{i=1}^{N} \left(\frac{p(x_i)}{1-p(x_i)}\right)^{y_i} (1 - p(x_i))^{ni} \quad (14)$$

Where:

$$Ln\left(\frac{p(x_i)}{1-p(x_i)}\right) = B_0 + B_1 x_i \quad (15)$$

By raising both sides of the equation to the natural base e, the equation becomes as follows:

$$\frac{p(x_i)}{1\text{-}p(x_i)} = e^{B_0 + B_1 x_j} \quad (16)$$

Substituting into equation (14) becomes:

$$L(B, Y) = \prod_{i=1}^{N} \left(e^{B_0 + B_1 x_j}\right)^{y_i} \left(1 - \frac{e^{B_0 + B_1 x_j}}{1 + e^{B_0 + B_1 x_j}}\right) \quad (17)$$

Taking the equation mentioned above:

$$Ln(L(B, Y) = \sum y_i (B_0 + B_1 x_j) - n_i Ln(1 + e^{B_0 + B_1 x_j}) \quad (18)$$

The following steps can be used to explain the steps of the LR algorithm:

Step 1: Determine the data matrix X of the original image to be identified.

Step 2: Segmentation of the original matrix data through the k-means algorithm and converting them into clusters, and here the conversion is made of only two groups (0,1), i.e., a group containing the number zero and the other group containing the number one to form the response variable Y.

Step 3: Estimating the parameters of the original image by generalizing the variable Y.

Step 4: Build a model of the original image entered and calculate its logarithm.

Step 5: The input image is compared with the rest of the images by the (contribution) of the original image parameters with the remaining image data to identify the corresponding image.

Step 6: The image with a lower Euclidean distance is the one that matches the original image when it comes to comparing it with the other images and the following formula illustrates the Euclidean distance:

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

2.4 Comparison Criteria:

To distinguish between various facial patterns, a set of criteria can be used. Some of these criteria are discussed below:

2.4.1 Accuracy rate:

This criterion is used to evaluate the performance of different methods, and the comparison between the best methods in recognition and its formula can be calculated as follows (Soni and Sahu, 2013):

$$\text{accuracy rate} = \frac{\text{Number of correctly recognized faces}}{\text{Total number of faces}} \times 100\% \quad (19)$$

2.4.2 Mean square error

The mean square error used to compare the models in the classification of facial patterns and can be calculated as follows (Huang and Song, 2022):

$$M.S.E = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 / n \quad (20)$$

3. Discussion of Results :

The dimensional reduction algorithms PCA and 2DPCA was used with the logistics regression LR to compare two experimental sets of data and different sample numbers (small, medium, large), i.e., number of images to be identified were chosen. To gain insights into the performance of the three methods under varying conditions to evaluate their performance in facial recognition, note that we have used the MATLAB programming language, the results of the two groups can be explained as follows:

P-ISSN 2518-5764
E-ISSN 2227-703X

1. ORL data set: For comparison between methods, different sample numbers (for small samples 4 individuals and 8 individuals were selected, with 10 different images per individual, so that the total number of images used in small samples became (40,80), and medium samples 14 individuals and 20 individuals were chosen, and by 10 different images per individual, so that the total number of images used in medium samples became (140,200) As for large samples, 26 individuals and 32 individuals were selected by 10 Different images for each individual, so that the total number of images used in large samples becomes (260,320).

The accuracy rate ACC as well as the mean square error MSE were calculated for the three methods for comparison between the best methods Table (1) shows the accuracy rate on the ORL data set.

Table 1: Comparison of the accuracy rate (ACC%) on the ORL data set

N / Methods	40	80	140	200	260	320
PCA	77.5	77	65	57	54.6154	42.8571
2DPCA	100	100	100	100	100	100
LR	100	100	100	100	100	100

We can observe, through the results of the table above, the efficiency of the logistic regression model LR and the two-dimensional principal component analysis 2DPCA in terms of high accuracy rate, which reached 100% and as shown in figure 6, their values are equal to all sample numbers and neither method is affected by increasing the sample number.

Figure 6: Results of correct recognition from the ORL dataset

While we note through the results of the principal component analysis PCA that this method is affected by the sample numbers, meaning when the sample numbers increase, the accuracy rate decreases in image recognition, unlike the rest of the other methods, which remained high accuracy despite the increase in sample numbers, which indicates the efficiency of both LR and 2DPCA methods compared with PCA.

P-ISSN 2518-5764
E-ISSN 2227-703X

Table 2: Comparison of the MSE on the ORL data set

N / Methods	40	80	140	200	260	320
PCA	7.59e+13	2.13e+14	2.73e+14	9.56e+14	3.23e+13	7.54e+12
2DPCA	2.1534e+06	1.6588e+06	1.5429e+06	1.7586e+06	2.0299e+06	1.5911e+06
LR	0.05180	0.16129	0.094786	0.2323	0.14534	0.0024414

Table (2) shows the results of the MSE values for the three methods on the ORL database, where we found that each method is affected differently from the other when increasing the sample numbers because each method has its behavior, there are methods in which the value of MSE increases by increasing the sample size, while other methods decreases when it increasing, this problem can be due to a number of possible reasons like, sampling variability, representational capacity , signal-to-noise ratio and computational complexity etc. Despite the image data being reduced to both PCA and 2DPCA, and varied sample numbers (small, medium, and large) being employed in comparison to other methods, we note the superiority of the logistical regression model LR by attaining the lowest value of MSE, this is followed by 2DPCA and then PCA.

2. Real data: Different sample numbers were selected to compare the methods, (small, medium, large), and for small samples (75,150) Images were selected medium samples were selected (300, 400) images, and for large samples were selected (800,600) images and the accuracy rate ACC was calculated as well as the mean squares error MSE for the three methods for comparison and table (3) shows the accuracy rate on real data using the logistic regression model LR with both PCA and 2DPCA.

Table 3: Comparison of the accuracy rate (ACC%) on the real data

N / Methods	75	150	300	400	600	800
PCA	100	100	100	100	100	100
2DPCA	100	100	100	100	100	100
LR	100	100	100	100	100	100

We can observe through the results of Table (3) above the high accuracy rate ACC which reached 100% and as shown in figure 7, for all three methods used, and equal to their values despite the different numbers of the samples used, and this indicates that not all methods are affected by the increase in the sample numbers, especially PCA. Where we note here its efficiency in real data in terms of high accuracy and for all sizes of samples, unlike its application to ORL data.

Figure 7: Results of correct recognition from the real data

Table 4: Comparison of the MSE on the real data

N / Methods	75	150	300	400	600	800
PCA	1.86e+16	2.94e+15	1.64e+16	2.55e+16	2.38e+15	1.18e+16
2DPCA	5.5476e+05	3.257e+05	8.8809e+05	3.3775e+06	9.5694e+05	7.9809e+05
LR	0.13063	0.19346	0.13267	0.13224	0.091936	0.1584

Table (4) shows the MSE values for the real data from both PCA and 2DPCA with the logistic regression model LR. We discovered that the effects of increasing sample numbers on each method differ, and even though the image data was reduced to both PCA and 2DPCA, we were able to observe the superiority of the LR logistic regression model by obtaining the lowest MSE value, and even when different sample numbers were used in comparison to other methods, followed by 2DPCA and then PCA.

4. Conclusion:

In this research, the algorithms of dimension reduction 2DPCA and PCA were used with the LR logistic regression model to compare their performance in facial recognition of the data used in this research. Using these techniques will simplify the learning process while improving the system's overall functionality, the face recognition system can more effectively deal with and achieve better classification accuracy for facial images with several images and different shots of the same person. we observe that both logistic regression and 2DPCA give a high rate of accuracy in all the cases of sample numbers on the ORL data set by giving 100% accuracy and still gives 100% on the real data. As for PCA, the value was affected when the sample numbers were different on the ORL data set, while PCA achieved a high rate of accuracy on the real data by giving 100%. The results also showed the efficiency of the LR logistic regression model in obtaining the lowest mean square error MSE for both sets of data despite the image data was reduced to both PCA and 2DPCA followed by 2DPCA and then PCA.

Authors Declaration:

Conflicts of Interest: None

-We Hereby Confirm That All The Figures and Tables In The Manuscript Are Mine and Ours. Besides, The Figures and Images, Which are Not Mine, Have Been Permitted Republication and Attached to The Manuscript.

- Ethical Clearance: The Research Was Approved By The Local Ethical Committee in The University.

References:

1. Abdulmunem, M. E., Ibrahim, F. B. (2016) " The Design of Efficient Algorithm for Face Recognition Based on Hybrid PCA-Wavelet Transform", Iraqi Journal of Science, Vol. 57, No.2A, pp. 995-1006.

2. Aggarwal, R., Bhardwaj, S., and Sharma, K. (2022) " Face Recognition System Using Image Enhancement with PCA and LDA", 6th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, pp. 1322-1327.

3. Al-Jubouri, A. AS. H.T. (2018) "Using SVM and LRM Logistic Regression Machine Model in Data Classification with Practical Application on Diabetes Patients at the General Ports Hospital in Basra", Master Thesis, University of Basra, Basra.

4. Al-Nasrawi, N.A.O. (2017) "Using the Bootstrap Method in the Analysis and Comparison of Parametric and Parametric Models", Master Thesis, University of Karbala, Karbala.

5. AL-Rawi, A.G., and Isa, A. M. (2019) "Use Principal Component Analysis Technique to Dimensionality Reduction to Multi-Source", Journal of Economics and Administrative Sciences, Vol. 25, No.115, pp. 464- 473.

6. Beiranvand, F., Mehrdad, V., and Dowlatshahi, M. B. (2022) " Unsupervised feature selection for image classification: A bipartite matching-based principal component analysis approach", Knowledge-Based Systems, Vol.250, No.109085, pp. 1- 14.

7. Bengherabi, M., Mezai, L., Harizi, F., Cheriet, M., and Guessoum, A. (2008) " Face recognition based on 2DPCA, DIAPCA and DIA2DPCA in DCT domain", 5th International Multi-Conference on Systems, Signals and Devices, Amman, Jordan, pp. 1-6.

8- Ebied, H. M. (2012). "Feature extraction using PCA and Kernel-PCA for face recognition", 8th International Conference on Informatics and Systems (INFOS) , pp. MM-72, IEEE.

9. Frangi, A. F., & Yang, J. Y. (2004) " Two-dimensional PCA: a new approach to appearance-based face representation and recognition ", IEEE transactions on pattern analysis and machine intelligence, Vol 26, No. 1, pp. 131-137.

10. Fırat, H., Asker, M. E., and Hanbay, D. (2022) "Classification of hyperspectral remote sensing images using different dimension reduction methods with 3D/2D CNN", Remote Sensing Applications: Society and Environment, Vol.25, No100694.

11. Huang, L., Song, T. (2022) " VLSI test through an improved LDA classification algorithm for test cost reduction", Microelectronics Journal, Vol. 125, No.105461, pp. 1-7.

12. Karthick, S., Selvakumarasamy, S., Arun, C., and Agrawal, P. (2021) "Automatic attendance monitoring system using facial recognition through feature-based methods (PCA, LDA) ", Journal Article in Materials Today: Proceedings, https://doi.org/10.1016/j.matpr.2021.01.

13. Khairunnahar, L., Hasib, M. A., Rezanur, R. H. B., Islam, M. R., and Hosain, M. K. (2019) "Classification of malignant and benign tissue with logistic regression", Informatics in Medicine Unlocked, Vol.16, No.100189, pp. 1-9.

14. Khalaf, N. B., and Mohammed, L. A. (2023) "Comparison of Some Methods for Estimating Nonparametric Binary Logistic Regression", Journal of Economics and Administrative Sciences, Vol. 29, No.135, pp.56-67.

15. Ma, J., Yuan, Y. (2019) " Dimension reduction of image deep feature using PCA", Journal of Visual Communication and Image Representation, Vol. 63, No.102578, pp. 1-8.

16. Pattnaik, I., Dev, A., and Mohapatra, A. K. (2023) " A face recognition taxonomy and review framework towards dimensionality, modality and feature quality", Engineering Applications of Artificial Intelligence, Vol.126, No.107056, pp. 1-26.

17. Prusty, M. R., Jayanthi, T., Chakraborty, J., and Velusamy, K. (2017) " Feasibility of ANFIS towards multiclass event classification in PFBR considering dimensionality reduction using PCA", Annals of Nuclear Energy, Vol. 99, No.12, pp. 311-320.

18. Saeed, R. A. (2015) "Using the Logistic Regression Model in Studding the Assistant Factors to Diagnose Bladder Cancer", Journal of Economics and Administrative Sciences.Vol.21, No. 83,pp.342-356.

19. Shanthi, K. G., Vidhya, S. S., Vishakha, K., Subiksha, S., Srija, K. K., and Mamtha, R. S. (2023). "Algorithms for face recognition drones", Materials Today: Proceedings, Vol.80, part. 3, pp.2224-2227.
20. Soni, S., Sahu, R.K. (2013) " Face recognition based on Two-Dimensional Principal Component Analysis (2DPCA) and Result in Comparison with Different Classifiers ", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 1, Issue 10, pp. 3899-3904.
21. Tharwat, A. (2016) "Principal component analysis-a tutorial", International Journal of Applied Pattern Recognition, Vol.3, No.3, pp.197-240.
22. The ORL Database, http://www.uk.research.att.com/facedatabase.html
23. Vanlalhruaia, V., Singh, Y. K., and Singh, N. D. (2017) "Binary face image recognition using logistic regression and neural network", International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), Chennai, India, pp. 3883-3888.
24. Wang, W., Qin, J., Zhang, Y., Deng, D., Yu, S., Zhang, Y., and Liu, Y. (2021) "TNNL: a novel image dimensionality reduction method for face image recognition", Digital Signal Processing, Vol.115, No.103082, pp. 1-11.
25. You, M., Cai, C. (2009)" Weed Seeds Classification Based on PCA, 2DPCA, Column-directional 2DPCA, and (2D) 2PCA", International Asia Symposium on Intelligent Interaction and Affective Computing, Wuhan, China, pp. 187-190.
26. Zhang, D., Zhou, Z. (2005) " $(2D)^2$ 2PCA: Two-directional two-dimensional PCA for efficient face representation and recognition ", Neurocomputing, Vol.69, No. (1-3), pp. 224-231.
27. Zhou, C., Wang, L., Zhang, Q., and Wei, X. (2014) "Face recognition based on PCA and logistic regression analysis", Optik, Vol.125, No.20, pp. 5916-5919.

Journal of Economics and Administrative Sciences
2024; 30(143), pp. 405-419
P-ISSN 2518-5764
E-ISSN 2227-703X

التعرف على صورة الوجه باستخدام بعض خوارزميات اختزال الابعاد والانحدار اللوجستي

سرى صباح كاتب
جامعة بغداد / كلية الإدارة والاقتصاد/قسم الإحصاء
العراق
sora.sabbah1101a@coadec.uobaghdad.edu.iq
https://orcid.org/0009-0007-1284-7662

انتصار عريبي فدعم
جامعة بغداد / كلية الإدارة والاقتصاد/قسم الإحصاء
العراق
entsar_arebe@coadec.uobaghdad.edu.iq
https://orcid.org/0009-0000-0316-6995سلطان

Received: 7/9/2023 Accepted:7/2/2024 Published Online First: 1 /10/ 2024

مستخلص البحث:

يعتبر التعرف على الوجه تقنية شائعة للذكاء الاصطناعي ومعالجة الصور ، حيث تتعرف خوارزميات التعلم الآلي تلقائيا على وجه الشخص من خلال مجموعة ضخمة من الصور أو حتى من مقطع فيديو مباشر. نظرا لوجود صعوبة في اكتشاف البيانات عند استخدام لقطات مختلفة لنفس الشخص في التعرف على الوجه ، يستغرق الامر وقتاً لتحديد عضوية الشخص .علاوة على ذلك عند العمل مع البيانات عالية الأبعاد حيث يمكن ان تشكل قيم البكسل ثنائية الابعاد لصور الوجه تحدياً حسابياً، وبالتالي تصبح مشكلة التعرف على الوجوه أكثر صعوبة ، لذا يقارن البحث الحالي تقنيات اختزال الأبعاد مثل خوارزمية تحليل المركبات الرئيسية PCA ، والتي هي خوارزمية تعلم غير خاضعة للإشراف تقلل الأبعاد عن طريق نقلها من منطقة عالية الأبعاد إلى منطقة ذات أبعاد أقل دون فقدان المعلومات المهمة ، وخوارزمية تحليل المركبات الرئيسية ثنائية الأبعاد 2DPCA لتحديد التقنية الأكثر فعالية لعملية اكتشاف صور الوجه و سيتم مقارنة هذه الطرق مع الانحدار اللوجستي وهو نموذج احصائي للتعرف على الوجوه تم استخدامه في مهام التصنيف. تم استخدام مجموعتين مختلفتين من البيانات تمثل المجموعة الأولى قاعدة بيانات ORL ، التي تتكون من 40 فردا ، بينما تمثل المجموعة الثانية البيانات الحقيقية ، والتي تتكون من 100 فرد . اظهرت النتائج ان نموذج الانحدار اللوجستي حقق أدنى متوسط مربعات الخطأ MSE لكلتا مجموعتي البيانات يليه 2DPCA و PCA ، وان التقنيات الثلاثة جميعها وصلت الى اعلى معدل دقة قدره 100% في التعرف على صور الوجه للبيانات الحقيقية.

نوع البحث: ورقة بحثية

المصطلحات الرئيسة للبحث: الانحدار اللوجستي, تحليل المركبات الرئيسية, تحليل المركبات الرئيسية ثنائية الابعاد, التعرف على الوجه, التعرف على الانماط.