

Comprehensive Study and Evaluation of Commonly used Dimensionality Reduction Techniques in Biometrics Field

Shahad A. Sultan
shahadali.sultan@gmail.com

M.F. Ghanim
mayada.faris@uomosul.edu.iq

Computer Engineering Department, Collage of Engineering, University of Mosul

Received: 7/7/2020

Accepted: 1/8/2020

ABSTRACT

In biometrics field, usually feature vectors have major length and contain ineffective information. This problem is so called "curse of dimensionality". Hence, there is a need for efficient dimensionality reduction technique to remove the redundant features and reduce the size of feature vectors to get high accuracy rate with fast performance. In this paper a comprehensive study of commonly used dimensionality reduction techniques: Principle Component Analysis, Linear Discremenant Analysis, and Generalized Discremenant Analysis, have been handled. Theoretical background of these techniques is illustrated along with the methods used to calculate their projection spaces then; practical implementation is conducted to find out and adopt the best one for retina based biometric authentication system. From this extensive study, it has been concluded that PCA technique has a number of problems make it has a bad classification power. LDA technique has a number of problems make it impossible to implement in most cases of biometrics field, while GDA technique is more efficient than the PCA and LDA techniques for dimensionality reduction purpose. It has high classification power and consumes less computational time. Hence, GDA technique is adopted in the proposed authentication system.

Keywords:

Biometric Authentication; Curse of Dimensionality; Feature Selection; PCA; LDA; GDA.

<https://rengj.mosuljournals.com>

Email: alrafidain_engjournal1@uomosul.edu.iq

1. INTRODUCTION

In biometrics, feature vectors usually suffer from the "curse of dimensionality" problem where redundant features in these vectors increase time complexity and degrade the performance of the authentication system. To solve this problem, there is a real need to remove ineffective features and transform dataset to a lower dimensional space [1][2][3].

Dimensionality reduction is used as a pre-processing step in many fields of machine learning related to the data mining; one of them is the biometrics field. Dimensionality reduction can be handled by using two techniques, feature selection and feature extraction which are explained as the following [2][4][5]:

- **Feature selection:** a subset of features **b** is selected from a given feature vector **T** based on features redundancy and relevance where, **b**

< **T**. feature selection is used to remove the noise, redundant and irrelevant features which results in improving the performance of the system [2][5]:

$$\begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_T \end{bmatrix} \xrightarrow{\text{Feature Selection}} \begin{bmatrix} x_{i1} \\ x_{i2} \\ \cdot \\ \cdot \\ x_{ib} \end{bmatrix} \quad (1)$$

- **Feature extraction:** it is based on features combinations, where it transforms the original features set, linearly or non-linearly, into a new feature set of lower dimension. The objective of this process is to produce more meaningful features in the new space [2][4]:

$$\begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_T \end{bmatrix} \xrightarrow{\text{Feature Extraction}} \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_b \end{bmatrix} = f \left(\begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_T \end{bmatrix} \right) \quad (2)$$

This research will concentrate on most commonly used feature extraction techniques: Principle Component Analysis (PCA) technique, Linear Discriminant Analysis (LDA) technique, and Generalized Discriminant Analysis (GDA) technique. A comprehensive study with practical implementation of these techniques has been presented. The main objective is to practically demonstrate the best one among them in order to adopt it for the dimensionality reduction and classification purposes in biometrics field.

Feature extraction algorithms can be categorized based on the presence or absence of class labels in the training set (learning method), as:

- **Supervised:** these algorithms are used when feature vectors in the training set (system database) have their associated class labels. For example, figure (1) shows a given training set that consists of five subjects (classes); each of them has three sample images. Each feature vector in the training set is associated with a certain class label; when using the supervised method these labels are used to find the lower dimensional space. LDA technique, GDA, and Mixture Discriminant Analysis (MDA) technique are examples of supervised algorithms [2][4][5][6].
- **Unsupervised:** these algorithms are used when feature vectors in the training set are not labeled. They find the lower dimensional space without using class labels, so they perform the most challenging task than the supervised

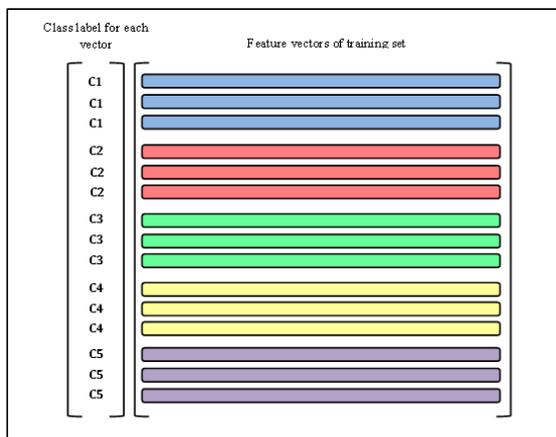


Fig. 1 Training set with its associated classes' labels

algorithms. PCA technique and Independent Component Analysis (ICA) technique are examples of unsupervised algorithms [2][4][5][6].

- **Semi-supervised:** this type of algorithms is used when the training set combines a small amount of labeled feature vectors and a large amount of unlabeled feature vectors. Semi-supervised algorithms fall between supervised algorithms (when all feature vectors are labeled) and unsupervised algorithms (when all feature vectors are unlabeled). They employ unlabeled data to assist the classification with more accuracy under situation of limited labeled data [2][7].

Also, feature extraction techniques can be categorized based on linearity as:

- **Linear algorithms:** these algorithms seek a linear transformation that sets apart different classes. However, if the classes are not linearly separable, linear algorithms fail to find a lower dimensional space where there will be a large overlap between the different classes [6][7][8]. Figure (2) illustrates an example of two linearly and two non-linearly separable classes.
- **Non-linear algorithms:** when classes are not linearly separable as shown in figure (2-b), non-linear feature extraction algorithms are used to seek a non-linear projection to discriminate among such classes [6][7][8].

The rest of this paper is organized as follows: Section 2 states the related works. Section 3 illustrates the dimensionality reduction techniques and all their related issues in detail. Section 4 handles the experimental results and discussion. The main concluded points are illustrated in Section 5.

2. RELATED WORKS

There is a number of research works in literature which have been made a study about the matter of dimensionality reduction techniques. For example, Sh. Wang et al. have presented in [9] several feature selection and feature extraction techniques for partial discharge pattern recognition. In order to compare the performance of these techniques, they carried out partial discharge tests on artificial partial discharge defect models. Alaa Tharwat in [5] has handled an extensive study in order to understand PCA technique. Then, she implemented PCA technique in the real applications. She has handled the same study on LDA technique in

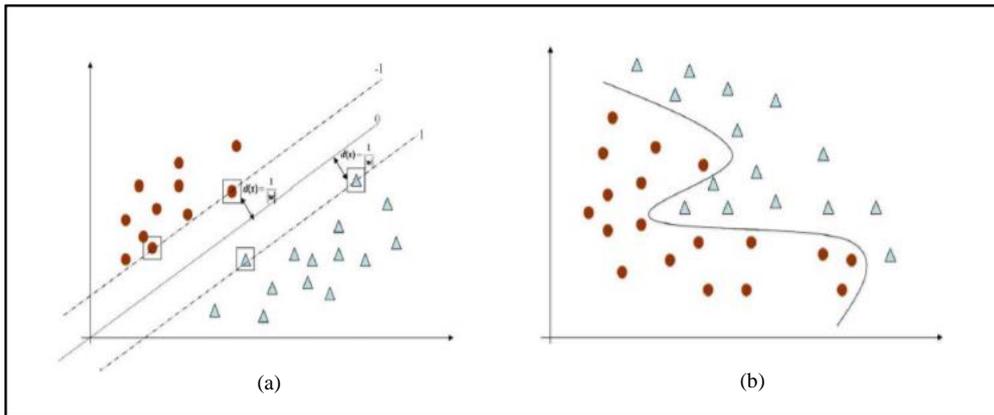


Fig. 2 Example of four classes [3] (a) Two linearly separable classes (b) Two non-linearly separable classes.

[6]. On the other hand, Vijaykumar N and I. Ahmed have presented in [3] a theoretical study on the most commonly used feature selection methods. Since the choice of feature selection techniques depends on the application areas, they have handled a comparative study on feature selection techniques based on several well known application domains. B. Venkatesh and J. Anuradha in [2] have presented a survey of feature selection techniques.

3. DIMENSIONALITY REDUCTION TECHNIQUES

Figure (3) shows the complete categorization of dimensionality reduction algorithms. Over the past few decades or so, a number of methods have been used to implement these algorithms. This research work will address three feature extraction methods: PCA, LDA, and

GDA which are the most commonly used in biometrics field.

3.1. Principle component analysis technique

PCA technique is one of the most famous linear unsupervised feature extraction algorithms. It seeks the space which represents the direction of the maximum variance of a given dataset [5][7][9][10][11].

PCA has a number of objectives involving: seeking relationships between samples, extracting the most important features from a given dataset, removing aberrant features, like noise which have a great impact on the classification process, and reducing the dimension of the dataset by retaining only important features.

These objectives can be achieved by

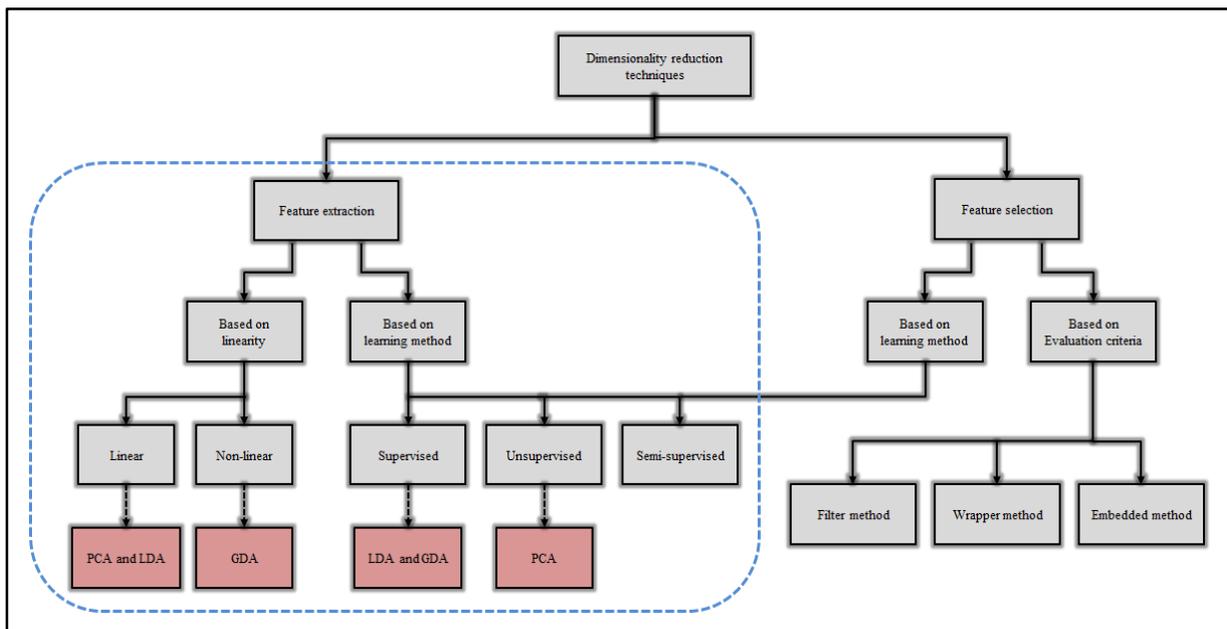


Fig. 3 Categorization of dimensionality reduction techniques

generating the PCA space [5][9]. PCA space is used to transform a given dataset to a lower dimensional space by projecting all samples of the dataset onto this lower space. PCA space consists of k orthogonal principle components (PCs). In this research work, covariance matrix method is used to calculate PCs [5]:

$$V\Sigma = \lambda V \quad (3)$$

Where, Σ is the covariance matrix of the dataset, V and λ are the eigenvectors and eigenvalues of the covariance matrix. Eigenvalues are scalar values and represent the robustness of the PCs. Eigenvectors are non-zero vectors which represents PCs themselves where, each eigenvector represents one principle component.

Principle components are uncorrelated and represent the direction of the maximum variance of the dataset. The first principle component (PC1 or v_1) represents the direction of the largest variance of the dataset, where the second principle component (PC2 or v_2) represents the direction of the second largest variance of the dataset, and so on. In other words, each principle component has a different robustness depending on the amount of variance in its direction. Usually, PCA space consists of the PCs that have the maximum amount of variance (maximum amount of the original data). In order to construct the PCA space (V_k), eigenvectors have to be sorted according to their corresponding eigenvalues. Then k eigenvectors that have the largest eigenvalues are selected as [5]:

$$V_k = [v_1, v_2, \dots, v_k] \quad (4)$$

Hence, increasing the number of the PCs in the PCA space will increase the robustness of the PCA technique which is measured as [5]:

Robustness of the PCA space =

$$\frac{\text{Total Variance of } V_k}{\text{Total Variance}} = \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^M \lambda_i} \quad (5)$$

Where, k is the number of selected eigenvectors and M is the total number of eigenvectors that are calculated from covariance matrix.

After constructing the V_k space, all samples of the dataset are projected onto this lower space as [5]:

$$Y = V_k^T R \quad (6)$$

Where, R is the mean centering samples of the dataset and Y is the resultant lower dimensional dataset.

3.1.1. Main problems of PCA technique

PCA technique suffers from a number of problems which make it not the best solution for "curse of dimensionality" problem in the biometrics field. These problems are [1]:

- **Linearity problem:** PCA technique is developed to generate a linear transformation for a given database to reduce its dimension. So, PCA technique is not suitable for non-linearly separable dataset [5][9].
- **Information Packing Transform problem:** The direction of the PCA projection space is determined by the maximum variance of a given dataset (i.e. maximum amount of original dataset). This direction may be useless for the classification process since it increases the total scatter across all classes in that dataset which leads to a bad class separability. Also, the PCA projection space in this direction may preserve useless information which degrades the system performance [1][12].
- **PCA technique does not care about classes of a given dataset:** it handles the overall dataset as a uniform matrix without concern about whether this dataset consists of one or more classes. It does not take the discrimination power into consideration.

Due to these problems, the PCA technique did not achieve satisfactory results when it was implemented in the proposed system as will be illustrated in section 4. So, there was a need to study the LDA technique to supersede the PCA technique.

3.2. Linear discriminant analysis technique

LDA technique is a very common linear supervised feature extraction algorithm. LDA transforms a given dataset into a lower dimensional space with more advantage than the PCA technique where the LDA projection space maximizes the ratio of the between-class variance to the within-class variance. This special advantage is very important in biometrics field for the classification process since it guarantees maximum class separability [8][9][10][11].

3.2.1. LDA algorithm

LDA lower dimensional space needs three main steps to be calculated. The first one is to find the separability between the different classes (the distance between the means of the different classes), which is represented by calculating the between-class matrix or variance [6][7]:

$$S_B = \sum_{j=1}^c E_j (\mu_j - \mu)(\mu_j - \mu)^T \quad (7)$$

Where, S_B is the between-class matrix; c is the number of classes in the dataset; E_j is the number of samples in the j^{th} class; μ_j is the mean of the j^{th} class; μ is the total mean of all samples in the dataset [6][7].

The second step is to find the distance between the mean of each class and its samples, which is represented by calculating the within-class matrix or variance [6][7]:

$$S_W = \sum_{j=1}^c \sum_{i=1}^{E_j} (I_{ij} - \mu_j)(I_{ij} - \mu_j)^T \quad (8)$$

Where, I_{ij} is the i^{th} sample in the j^{th} class.

The final step is to construct the projection space which can maximize the ratio of the between-class variance to the within-class variance. The transformation matrix of the LDA technique (W) is calculated as [6][7]:

$$W = S_W^{-1} S_B \quad (9)$$

Then calculate the eigenvalues ($\lambda = \{\lambda_1, \lambda_2, \dots, \lambda_M\}$) and eigenvectors ($V = \{v_1, v_2, \dots, v_M\}$) of equation (9) providing S_W is non-singular¹.

As with the PCA technique, eigenvectors represent the direction of LDA space where each eigenvector represents one axis of the new space. Also, the corresponding eigenvalues represent the robustness of these eigenvectors. Robustness of an eigenvector reflects its ability to discriminate among different classes by increasing the between-class variance and decreasing the within-class variance. Hence, eigenvectors have to be sorted in descending order depending on their corresponding eigenvalues. Then, the first k eigenvectors are selected to construct the LDA lower dimensional space (V_k).

To obtain a lower dimensional dataset, project all samples of a given dataset (X) onto the V_k space:

$$Y = X V_k \quad (10)$$

¹ A matrix is considered a singular matrix when it is square and does not have a matrix inverse, the determinant is zero; hence, not all columns and rows are independent.

3.2.2. Main problems of LDA technique

Although the LDA technique is one of the most commonly used feature extraction algorithms, it suffers from two essential problems: the linearity problem and the Small Sample Size (SSS) problem. In this section each problem is illustrated in detail with its some state-of-the-art solutions [6][7]:

- **Linearity problem:** As with the PCA technique, LDA is a linear algorithm so, if the dataset is non-linearly separable, the LDA fails to find a lower dimensional space. This means that the LDA technique fails when discriminatory information does not exist in the means of classes but in the variance. When the means are approximately equal, S_B and then W will be zero so, there is no LDA space for that dataset [6][7].

One of the common solutions for this problem is by using kernel methods or functions. Kernel functions transform the original dataset into a higher dimensional space which is linearly separable.

- **Small Sample Size problem:** Small Sample Size (SSS) problem is also known as Singularity or Under-Sampled problem. It is one of the great problems of the LDA technique especially when used in biometrics field. It results from high dimensional feature classification tasks or when there is a small number of samples available within each class compared to the number of features extracted from each sample (dimension of the dataset). This causes S_w to be a singular matrix where the upper limit of the rank² of S_w is $N-c$, while the dimension of S_w is $(M \times M)$. In biometrics field, in most cases $M \gg (N-c)$ which results in SSS problem. There are many common solutions for this problem; each has its advantages and drawbacks [6][7]:

- **Regularization:** This solution adds the identity matrix, after multiplying it by a regularization parameter ($\eta > 0$), to the S_w to make it non-singular and find its inverse. So, diagonal elements of S_w are biased as: $S_w = S_w + \eta I$. This method requires more tuning to choose an appropriate value of regularization parameter. Also, incorrect choice of this value causes degradation for the performance of the system. A variant of LDA technique called Regularized

² The rank of the matrix represents the maximum number of linearly independent rows or the maximum number of linearly independent columns in the matrix.

LDA (RLDA) uses this method to eliminate the SSS problem [6][7].

- **Sub-space:** this solution uses an intermediate method to reduce the dimension of a given dataset to be equal to the rank of SW; so, SW becomes a full-rank (invertible). For example, PCA is used to reduce the dimension of the original dataset to be equal to (N-c) the upper limit of the rank of SW. Then, LDA is used to further reduce features to k-dimensions without causing the SSS problem. This method causes the loss of some discriminant information which results in the degradation of the system performance. A variant of LDA technique called PCA+LDA technique uses this solution [6][7].
- **Null space:** This method removes the null space of SW to make it full-rank; thereby invertible. When the null space is removed from SW, more discriminant information is lost then affecting the system performance. Null LDA technique, (NLDA) technique, uses null space method to eliminate the SSS problem [6][7].

Figure (4) shows the difference between LDA technique and PCA technique in terms of the used mechanism to construct the lower dimensional projection space. Features in this figure are extracted from phase resolved partial discharge pattern and partial discharge waveforms to represent and recognize typical defects. The LDA shows maximum separation

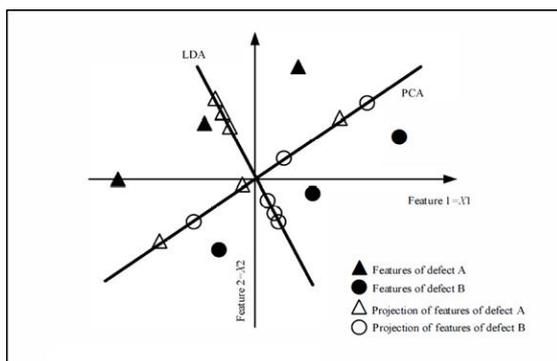


Fig. 4 The difference between PCA and LDA projection space [9].

between two classes (defect A and defect B) which leads to better performance than the PCA technique [9]. However, the main problems of LDA technique make it impossible to be implemented in the proposed

system. So, there has been a need to study GDA technique.

3.3. Generalized discriminant analysis technique

Generalized Discriminant Analysis or Kernel Discriminant Analysis (GDA) is a non-linear supervised feature extraction technique. GDA is a kernel version of LDA; it is the more general case and used in this research work to eliminate any shortcomings of both the PCA and LDA techniques. Similar to LDA, GDA seeks a projection space that transforms features into a lower dimensional space and maximizes the ratio of the between-class variance to the within-class variance. With the GDA space the most valuable information is preserved which indicates high classification efficiency and reduces the training time of the used classifier [1][7][9][13].

3.3.1. Calculating GDA projection space

The GDA projection space is calculated as the following [1][13]:

To eliminate the linearity problem, GDA is based on a kernel function ϕ which transforms the original dataset X into a higher dimensional space Z where:

$$\phi: X \rightarrow Z \tag{11}$$

Then calculate the between-class matrix of the non-linearly mapped data (in Z space):

$$S_B^\phi = \sum_{j=1}^c E_j (\mu_j^\phi - \mu^\phi)(\mu_j^\phi - \mu^\phi)^T \tag{12}$$

Where, S_B^ϕ represents the between-class matrix in Z space; μ_j^ϕ represents the mean of j^{th} class in Z space; μ^ϕ represents the total mean of the dataset in Z space;

The within-class matrix in Z space is calculated as:

$$S_W^\phi = \sum_{j=1}^c \sum_{i=1}^{E_j} (\phi(I_{ij}) - \mu_j^\phi)(\phi(I_{ij}) - \mu_j^\phi)^T \tag{13}$$

Where, S_W^ϕ represents the within-class matrix in Z space. The transformation matrix of the GDA technique (W) is calculated as:

$$W^\phi = \sum_{j=1}^c \sum_{i=1}^{E_j} \alpha_{ji} \phi(I_{ji} - \mu_j^\phi) \tag{14}$$

α_{ji} represents a vector of some real weights. Eigenvalues are calculated as:

$$\lambda = \frac{\alpha^T ADA \alpha}{\alpha^T AA \alpha} \tag{15}$$

Where, A represents the kernel matrix, it is of $(M \times M)$ dimension. If A is not reversible then the regularization process is used to eliminate the SSS problem. D represents a $(M \times M)$ block diagonal matrix.

At this point, determine the k eigenvectors which have the largest eigenvalues to construct the GDA projection space V_k . Then, the lower dimensional dataset is calculated as:

$$Y = A_x V_k \quad (16)$$

A_x represents the mapped dataset (using kernel function).

As illustrated above, GDA technique transforms a given dataset into a higher dimensional space using kernel function to make its classes are linearly separable. The same steps of the LDA technique are then applied to the mapped dataset to reduce its dimension. It selects those eigenvectors which have best classification capability than those eigenvectors which best describe the dataset (as with PCA technique) [1][7][9][12][13]. Hence, it can be said that the GDA technique can eliminate the problems of the PCA (linearity and poor discrimination capability) and the problems of the LDA (linearity and SSS problem).

4. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, the dimensionality reduction techniques which are mentioned in section 3 are implemented in a retinal-based identification system. This system has been designed as shown in figure (5). The objective of the following experiments is to assess the best technique among PCA, LDA, and GDA for the mentioned identification system.

4.1. Preparing the environment of experiments

All experiments in this section are conducted in the same environment which is composed of: Windows 10 Pro operating system, Intel (R) Core (TM) CPU @ 1.8 GHz, 8 GB RAM, and Matlab (R2019b). Also; these experiments are performed using the following databases:

- Retinal Identification DataBase (RIDB), which was designed by J. Fatima, A. M. Syed, and M. U. Akram during their research work in reference [14]. RIDB database contains 100 images of resolution 1504×1000 (collected from 20 different individuals with 5 images per individual).

In RIDB database, each user has five images, so three of them were used to train the system and the remaining two images were used for testing. Moreover, 17 out of 20 individuals were chosen as registered users in the system dataset and 3 individuals were not registered in the dataset and were considered as intruders to the system. Hence, the system training dataset consisted of 51 retinal images (17 individuals \times 3 training images). Whereas the system performance was tested using 34 retinal images as genuine users (17 individuals \times 2 testing images) and 15 retinal images as impostors (3 unregistered individuals \times 5 images).

- Digital Retinal Images for Vessel Extraction (DRIVE) database. This database was acquired in the Netherlands from a diabetic retinopathy checking program. Checking people consisted of 400 diabetic subjects between 25-90 years old. Forty images of resolution 768×584 pixels have been randomly opted from them to construct the online database.

DRIVE database is rotated based on the "Data Augmentation" concept [15][16]. Rotation angles applied to retinal images are: $\pm 10^\circ$, $\pm 15^\circ$, $\pm 20^\circ$, $\pm 25^\circ$, $\pm 30^\circ$, $\pm 35^\circ$. After these rotation processes, the number of images becomes 500.

For DRIVE database, 34 out of 40 individuals were chosen as registered users and 6 individuals were considered as intruders. The system training dataset consisted of 170 retinal images (34 individuals \times 5 training images), whereas the system was tested using 272 retinal images as genuine users (34 individuals \times 8 testing images) and 58 retinal images as impostors.

There is a number of criteria used to evaluate the performance of biometric authentication systems [14][17][18][19]. Some of them are illustrated below:

- **False Acceptance Rate (FAR):** FAR is the ratio of the subjects which are incorrectly accepted.

$$FAR = \frac{\text{number of accepted imposters}}{\text{number of imposter comparisons}} \times 100\% \quad (35)$$

- **False Rejection Rate (FRR):** FRR is the ratio of subjects which are incorrectly rejected.

$$FRR = \frac{\text{number of rejected genuine users}}{\text{number of genuine comparisons}} \times 100\% \quad (36)$$

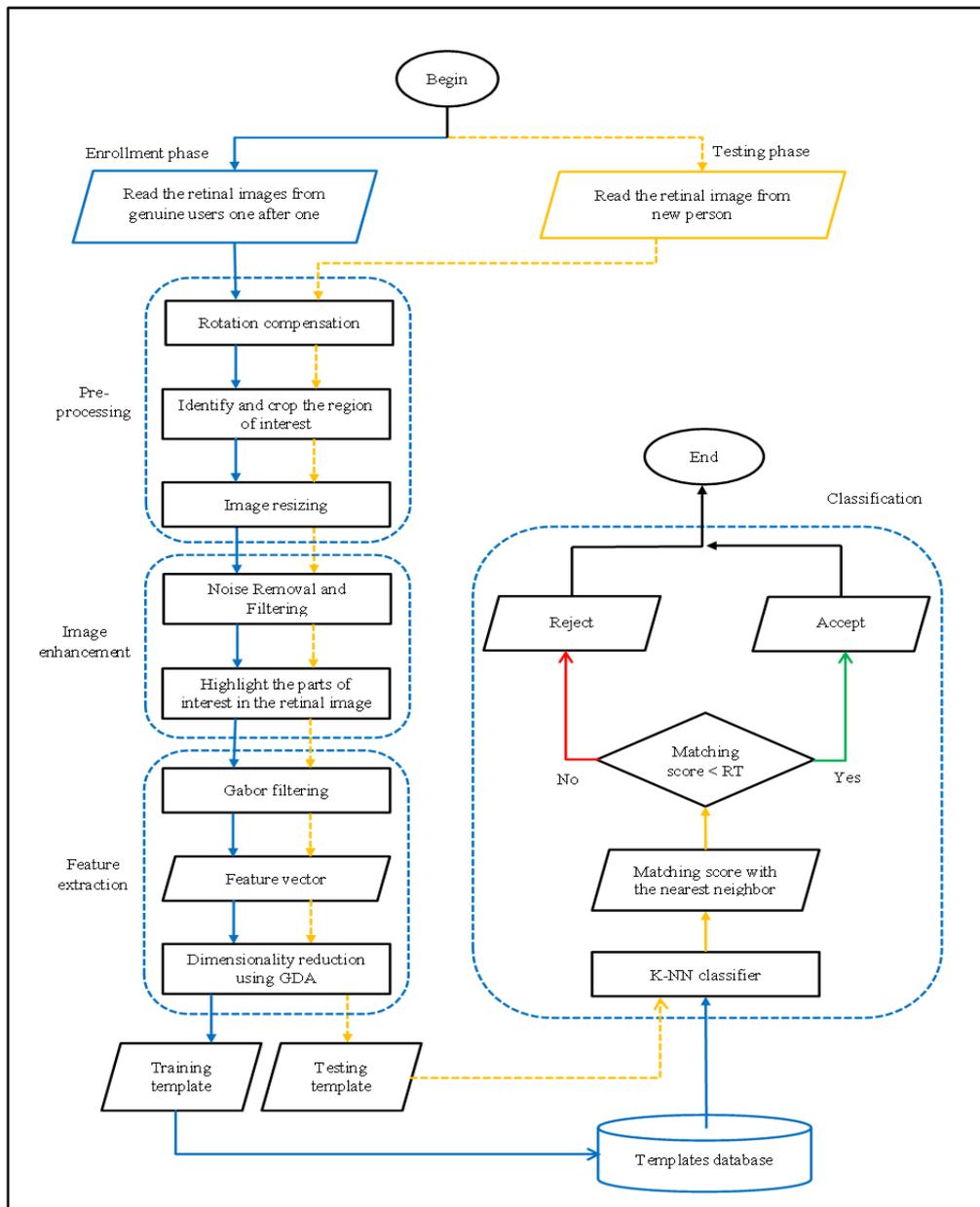


Fig. 5 Proposed algorithm of retinal based identification system

- Receiver Operating Characteristic (ROC):** ROC is the curve that represents the relation between FAR and FRR. It represents a function of threshold value and abstracts the performance of the biometric authentication system.
- Equal Error Rate (EER):** EER is the rate at which both FAR and FRR having equal values. EER can easily be found out from the ROC curve. The most accurate system has the lowest EER value. EER can be computed at the point where:

$$FAR(t) = FRR(t) \tag{37}$$

But practically, the distribution of the matching scores (threshold values) is not continuous and the crossover point may not exist within these distributions. In this case EER can be calculated as:

$$EER = \begin{cases} \frac{FAR(t_1)+FRR(t_1)}{2} & \text{if } FRR(t_1) - FAR(t_1) \leq FAR(t_2) - FRR(t_2) \\ \frac{FAR(t_2)+FRR(t_2)}{2} & \text{otherwise} \end{cases} \tag{38}$$

$$t_1 = \max_{t \in g} \{t | FRR(t) \geq FAR(t)\} \quad (38-a)$$

$$t_2 = \min_{t \in g} \{t | FRR(t) \leq FAR(t)\} \quad (38-b)$$

g Represents the set of reference threshold values used in the score distributions

- **Accuracy:** It is the rate of individuals that are correctly classified, accuracy of a given system is calculated as:

$$\text{Accuracy (\%)} = \left[100 - \left(\frac{FAR(\%) + FRR(\%)}{2} \right) \right] \quad (39)$$

So, accuracy of the system improves if the values of FAR and FRR decreases.

4.2. Implementation of PCA technique

In this section we will investigate the effect of the number of selected PCs (k) to construct the PCA space, on the system performance. RIDB database is used to implement PCA technique. Computational time and accuracy rate are considered as evaluation criteria for experiment results. In this experiment, different numbers of PCs are used to construct the PCA space. As a result, the dimension of the projected training set and testing vector are changed based on the value of k . Figure (6) illustrates the results of this experiment. This figure shows that computational time and accuracy rate of the system are proportional to the number of the selected eigenvectors. Hence, when using PCA technique, the trade-off between these parameters should be considered. As mentioned earlier, eigenvectors in the PCA space are sorted according to their robustness, where robustness of each eigenvector reflects its ability to discriminate among different classes. This means that increasing the number of

eigenvectors preserves more important information in the projected feature vectors. Hence, as shown in figure (6-a), when increasing the number of eigenvectors from 0.1% to 50% of the total number, the identification accuracy increases from 39.6% to 85.3%. Also, increasing the number of selected eigenvectors increases the dimension of PCA space and thus the dimension of the projected feature vectors. This causes the computational time to be increased from 3.3 to 3.6 seconds as shown in figure (6-b).

When the number of the selected eigenvector is 50% of the total number, the size of the feature vector becomes 8000 in RIDB database and 4500 in DRIVE database, which is considered a long vector. So, at this point increasing process is stopped since more increasing will require more computational time which is inconsistent with the real time objective of this research work. Also, the remaining eigenvectors which are to be selected have less ability for classification than those of first selection and it is not expected to considerably increase the identification accuracy.

4.3. Implementation of LDA technique

In biometrics field and especially in this research work, it is impossible to implement LDA technique as a dimensionality reduction method due to its linearity and the SSS problems. Features extracted from retinal images are non-linearly separable also; in biometrics field usually the system database has a number of samples for each subject less than the dimension of each sample (SSS problem). As mentioned earlier there is a number of the LDA variants used to eliminate the LDA problems; GDA is considered one of them.

4.4. Implementation of GDA technique

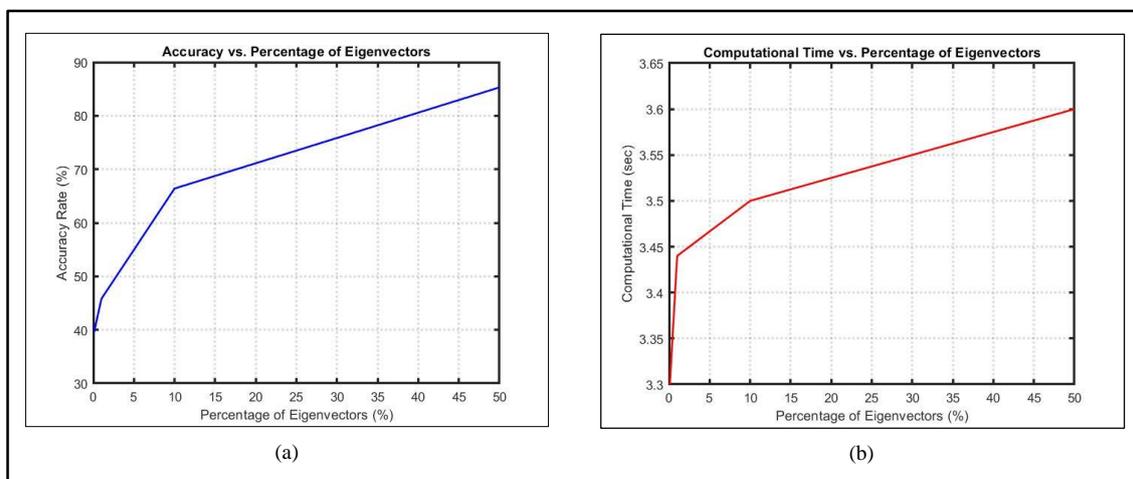


Fig. 6 Accuracy and CPU time as a function of different numbers of eigenvectors in the PCA space

Using GDA technique, the maximum number of eigenvectors which can be selected to construct the GDA space is (number of classes – 1). So, it can considerably reduce the dimension and preserve the most important information due to its high classification power. Results of implementing GDA technique for both RIDB database and DRIVE database are shown in figures (7, 8, 9 and 10). Figure (7) shows that the proposed system is not sensitive to threshold values in the range between 24 and 34. In this region, FAR = FRR = ERR = zero, which represents the ideal performance for high security level application.

Hence, the OP of proposed system is determined to be here by making the threshold value of the classification process in the range of [24 to 34]. Figure (8) also demonstrates the excellent performance of the proposed system. It shows a good separation distance between genuine and imposter classes. Figures (9 and 10) are related to DRIVE database and also demonstrate the same excellent performance. Table (1) illustrates a comparison between PCA and GDA performance where GDA technique extremely outperforms PCA technique.

5. CONCLUSION

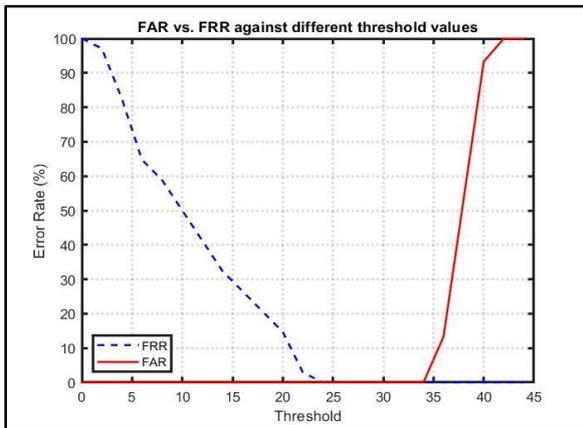


Fig. 7 FAR and FRR as functions of threshold values using RIDB database

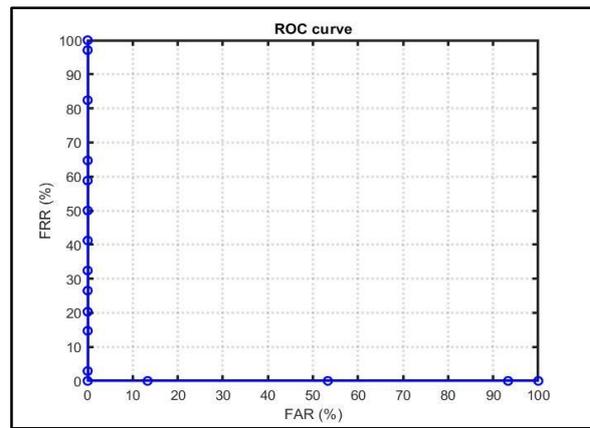


Fig. 8 ROC curve for RIDB database

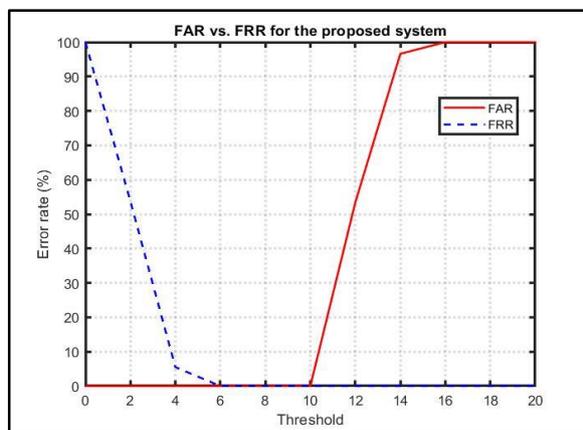


Fig. 9 FAR and FRR as functions of threshold values using DRIVE database

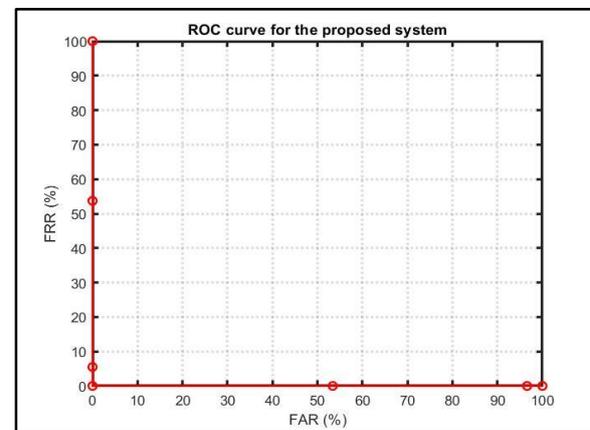


Fig.10 ROC curve for DRIVE database

TABLE (1) COMPARISON BETWEEN PCA AND GDA DIMENSIONALITY REDUCTION TECHNIQUES

Dimensionality reduction technique	Used database	Length of feature vector	Computational time (sec)	ERR (%)	Accuracy rate (%)
PCA	RIDB	8000	3.6	14.8	85.3
	DRIVE	4500	1.53	19.7	80.3
GDA	RIDB	16	0.22	0	100
	DRIVE	16	1.2	0	100

This research practically demonstrates that GDA technique is more suitable than PCA and LDA techniques for dimensionality reduction purpose in biometric field. Also, it consumes a less computational time. The PCA has a number of problems (linearity problem, information packing transform problem, and PCA technique does not care about classes of a given dataset) which make it has a bad classification power. The LDA has a number of problems (linearity problem and small sample size problem) which make it impossible to implement in most cases of biometrics field.

REFERENCES

- [1] M. Haghghat, S. Zonouz, M. Abdel-Mottaleb, "CloudID: Trustworthy cloud-based and cross-enterprise biometric identification", *Expert Systems with Applications*, Vol. 42, Issue: 21, pp. 7905–7916, 2015, <http://dx.doi.org/10.1016/j.eswa.2015.06.025>.
- [2] B. Venkatesh, J. Anuradha, "A Review of Feature Selection and Its Methods", *Cybernetics and Information Technologies*, Vol.19, Issue: 1, pp. 3-26, 2019, DOI: 10.2478/cait-2019-0001.
- [3] N. Vijaykumar, I. Ahmed, "Multimodal Biometric System Using Particle Swarm Based Feature Selection", *2017 International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies (ICAMMAET)*, Chennai, India. 16-18 Feb. 2017, pp. 1-6, 2017, DOI: 10.1109/ICAMMAET.2017.8186710.
- [4] A. Jović, K. Brkić, N. Bogunović, "A review of feature selection methods with applications", *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. Opatija, Croatia. 25-29 May 2015, pp. 1200-1205, 2015, DOI: 10.1109/MIPRO.2015.7160458.
- [5] A. Tharwat, "Principal Component Analysis - A Tutorial", *Int. J. Applied Pattern Recognition*, Vol.3, Issue: 3, pp. 197-240, 2016, DOI:10.1504/ijapr.2016.079733.
- [6] A. Tharwat, "Linear discriminant analysis: A detailed tutorial", *AI Communications*, Vol.30, Issue: 2, pp.169–190, 2017, DOI 10.3233/AIC-170729.
- [7] Y. Zhang, D. Yeung, "Semi-Supervised Generalized Discriminant Analysis", *IEEE Transactions on Neural Networks*. Vol.22, Issue: 8, pp. 1-11, 30 Jun. 2011, DOI: 10.1109/TNN.2011.2156808.
- [8] O. Ayad, M. Syed-Mouchaweh, "Multiple Classifiers Approach based on Dynamic Selection to Maximize Classification Performance", *International Journal of Machine Learning and Computing*, Jun. Vol.1, Issue: 2, pp. 154-162, 2011, DOI: 10.7763/IJMLC.2011.V1.23.
- [9] Sh. Wang, J. Zhang, H. Hu, Q. Liu, M. Zhu, H. Mu, G. Zhang. "Feature Optimization Selection and Dimension Reduction for Partial Discharge Pattern recognition", *2016 International Conference on Condition Monitoring and Diagnosis*, Xi'an, China, pp. 877-880, 2016. DOI: 10.1109/CMD.2016.7757962.
- [10] F. Feng , W. Li, Q. Du, B. Zhang, "Dimensionality Reduction of Hyperspectral Image with Graph-Based Discriminant Analysis Considering Spectral Similarity", *Remote sensing*. Vol.9, Issue: 4, pp. 1-12, 2017, DOI: 10.3390/rs9040323.
- [11] S. H. Cha, T. Cha, "Variance Linear Discriminant Analysis for IRIS Biometrics", *The Thirty-Third International FLAIRS Conference (FLAIRS-33)-Association for the advancement of Artificial Intelligence*, USA, pp. 504-507, 2020.
- [12] A. Khan, H. Farooq, "Principal Component Analysis-Linear Discriminant Analysis Feature Extractor for Pattern Recognition", *IJCSI International Journal of Computer Science Issues*. Nov. Vol.8, Issue: 6, pp. 267-270, 2011.
- [13] Sh. Singh, S. Silakari, R. Patel, "An efficient feature reduction technique for intrusion detection system", *2009 International Conference on Machine Learning and Computing*. Singapore, pp. 147-153, 2011.
- [14] J. Fatima, A. M. Syed, M. U. Akram, "A secure personal identification system based on human retina", *2013 IEEE Symposium on Industrial Electronics and Applications (ISIEA)*, Kuching, pp. 90–95, 2013, DOI: 10.1109/ISIEA.2013.6738974.
- [15] L. Taylor, G. Nitschke, "Improving Deep Learning using Generic Data Augmentation", *arXiv archive*, Eprint: 1708.06020, pp. 1-6, 20 Aug. 2017.
- [16] R. Khokher, R. Ch. Singh, A. Jain, "Verification of Biometric Traits using Deep Learning", *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*. Vol.8, Issue: 10S, pp. 452-459, Aug. 2019, DOI: 10.35940/ijitee.J1083.08810S19.
- [17] S. Shankar, V. R. Udupi, R. D. Gavas, "Biometric Verification, Security Concerns and Related Issues", *I.J. Information Technology and Computer Science*, pp. 42-51, 2016, DOI: 10.5815/ijitcs.2016.04.06.
- [18] J. Malik, Dh. Girdhar, R. Dahiya, G. Sainarayanan, "Reference Threshold Calculation for Biometric Authentication", *I.J. Image, Graphics and Signal Processing*, Vol.6, Issue: 2, pp. 46-53, 2014, DOI: 10.5815/ijigsp.2014.02.06.
- [19] A. J. Mansfield, J. L. Wayman/ D. Rayner, "Best Practices in Testing and Reporting Performance of Biometric". *National Physical Laboratory*, 14/02, 2002.

دراسة وتقييم شامل لتقنيات تقليل الأبعاد شائعة الاستخدام في مجال القياسات الحيوية

ميادة فارس غانم
mayada.faris@uomosul.edu.iq

شهد علي سلطان
shahadali.sultan@gmail.com

جامعة الموصل - كلية الهندسة - قسم هندسة الحاسوب

الملخص

في مجال القياسات الحيوية، عادةً يكون طول متجه الخصائص كبير جداً ويحتوي على معلومات غير مجدية. هذه المشكلة تعرف بـ "لعنة الأبعاد". بالتالي هناك حاجة لتقنية كفاءة تعمل على تقليل الأبعاد بإزالة الخصائص الزائدة عن الحاجة وتقليل حجم متجه الخصائص لغرض الحصول على معدل دقة عالي مع أداء سريع. تتناول هذه الورقة البحثية دراسة شاملة لتقنيات تقليل الأبعاد شائعة الاستخدام وهي: تقنية تحليل المكونات الأساسية، تقنية تحليل التمييز الخطي، وتقنية تحليل التمييز المعمم. يتم توضيح الخلفية النظرية لهذه التقنيات بالإضافة إلى الطرق المستخدمة لحساب فضاء الإسقاط الخاص بهم؛ ومن ثم إجراء التنفيذ العملي لغرض إيجاد واعتماد أفضلها في بناء نظام مصادقة بيومترية قائم على أساس شبكية العين البشرية. من هذه الدراسة الشاملة تم الاستنتاج بأن تقنية تحليل المكونات الأساسية لها عدد من المشاكل تجعلها ذات قدرة تصنيف رديئة. تقنية تحليل التمييز الخطي أيضاً تمتلك عدد من المشاكل تجعل من المستحيل تنفيذها في معظم حالات مجال القياسات الحيوية. أما تقنية تحليل التمييز المعمم فلها قدرة تصنيف عالية جداً وتستهلك وقت حسابي أقل لذا فهي أكثر كفاءة من تقنية تحليل المكونات الأساسية وتقنية تحليل التمييز الخطي لغرض الاستخدام في مجال تقليل الأبعاد. نتيجة لذلك تم اعتماد تقنية تحليل التمييز المعمم في نظام المصادقة المقترح.

الكلمات الدالة :

المصادقة البيومترية، لعنة الأبعاد، استخراج الخصائص، تحليل المكونات الأساسية، تحليل التمييز الخطي، تحليل التمييز المعمم.