Original Research Article

A Comparative Analysis of Selected Fisher Linear Discriminant Based Algorithms in Human Faces

.

ABSTRACT

One of the most reliable biometrics when issues of access control and security is been considered is face recognition. An integral part of a face recognition system is the feature extraction stage, which becomes a critical problem where is a need to obtain the best feature with minimum classification error and low running time. Many of the existing face recognition systems have adopted different linear discriminant-based algorithms independently for feature extraction in which excellent performance were achieved, but identifying the best most suitable of these variants of linear discriminant-based algorithms for face recognition systems remains a subject open for research. Therefore, this paper carried out a comparative analysis of the performance of the basic Linear Discriminant Algorithm (LDA) and two of its variants which are Kernel Linear Discriminant Analysis (KLDA) and Multiclass Linear Discriminant Analysis (MLDA) in face recognition application for access control.

Three Hundred and forty (340) face images were locally acquired with default size of 1200 x 1200. Two hundred and forty (240) images were used for training while the remaining hundred (100) images were used for testing purpose. The image enhancement involves converting into grayscale and normalizing the acquired images using histogram equalization method. Feature extraction and dimension reduction of the images were done using each of LDA, KLDA and MLDA algorithms individually. The extracted feature subsets of the images from each of LDA, KLDA and MLDA algorithm were individually classified using Euclidian distance. This technique was implemented using Matrix Laboratory (R2015a). The performance of LDA, KLDA and MLDA was evaluated and compared at 200 x 200 pixel resolution and 0.57 threshold value using recognition accuracy, sensitivity, specificity, false positive rate, training time and recognition time.

The evaluation result shows that the LDA algorithm yielded recognition accuracy, sensitivity, specificity, false positive rate, training time and recognition time of 93.00%, 92.86%, 93.33%, 6.67%, 1311.76 seconds and 67.98 seconds respectively. Also, KLDA recorded recognition accuracy, sensitivity, specificity, false positive rate, training time and recognition time of 95.00%, 95.71%, 93.33%, 6.67%, 1393.24 seconds and 63.67 seconds respectively. Furthermore, MLDA algorithm yielded recognition accuracy, sensitivity, specificity, false positive rate, training time and recognition time of 97.00%, 97.14%, 96.67%, 3.33%, 1191.55 seconds and 58.65 seconds respectively. The t-test measured between the accuracies of MLDA algorithm and KLDA reveals that MLDA algorithm was statistically significant at

$P < 0.05; P = 0.014\ and\ \mu = 1.50$. Also, the t-test measured between the accuracies of MLDA algorithm and LDA reveals that MLDA algorithm was statistically significant at $P < 0.01; P = 0.001\ and\ \mu = 3.75$.

Keywords: Biometrics, Face, Feature extraction, LDA, KLDA, MLDA.

1. INTRODUCTION

With continuous increase in world population, identification and authentication of individuals is becoming more significantly important. Hence, the need for highly accurate, secured and practical identification and authentication systems. Over the years, many traditional identification and authentication systems such as usernames, passwords, keys, personal identification number (PIN), identification (ID) cards, hardware token- based systems have been use for access control, but each of them has its own attendant problems. Generally, they are not reliable and secure in many of the security zones. Thus, there is an increasing need for an automatic and reliable identification and authentication systems. Biometric identification has proven to be more reliable means of verifying the human identity [27]. Biometrics is the science of establishing human identity by using physical or behavioral traits such as face, fingerprints, palm prints, iris, hand geometry and voice [28]. The work focuses on face recognition as a form of biometric identification and authentication technique.

Face recognition is a technology which recognizes human by his/her face image. Face recognition has attracted much attention and is still attracting the interest of many researchers in the area of pattern recognition, machine learning, and computer vision because of its immense application potentials [17]. Generally, facial recognition involves four major stages. These stages include image acquisition, image pre-processing, feature extraction and image classification. Of these four major stages, feature extraction is the most essential. Basically, it consists of extracting the most relevant features of an image and assigning it into a label [19]. Extracting features from face images for detection and recognition purpose is a central issue for face recognition systems [5]. Although feature extraction methods provide researchers with the main features that are associated with the face image sufficient enough to make good recognition, the feature set produced by these methods have very large dimension [4]. Hence, the need for dimensionality reduction. Dimensionality reduction plays crucial role in the face recognition problem. It is generally applied for improving robustness and reducing computational complexity of the face recognition problem. Out of all approaches available at hand, those based on appearance is considered to be most favourable. Therefore, methods like Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are used for dimensionality reduction and hence can provide efficient matching of features of faces for recognition purposes [30].

Furthermore, extracting proper features is crucial for satisfactory design of any pattern classifier, and how to develop a general procedure for effective feature extraction remains an interesting and challenging problem [14]. Traditionally, PCA has been the standard approach to reduce the high-dimensional original pattern vector space into low-dimensional feature vector space. Comparative studies between Fisher Linear Discriminant Analysis (FLDA) and Principal Component Analysis (PCA) on the face recognition problem were reported independently by [6] and [12], in which FLDA out performed PCA significantly. These successful applications of FLDA have drawn a lot of attention on this subject and the ensuing years have witnessed a burst of research activities on various issues relating to applying subspace methods such as PCA and FLDA to pattern recognition problems, with the latest development being an attempt to unify all these subspace methods under the same framework [18]. LDA provides fast feature extraction and classification due to its discriminative power and computational simplicity. Variants of LDA include LDA, Kernel-LDA

(KLDA), Incremental LDA (ILDA) and Multiclass LDA (MLDA) [34]. They have been widely applied in many applications of pattern recognition, computer vision, face recognition, text-image combination multimedia retrieval, speech and music classification, outliers detection, generalized image and video classification and so on [9].

This paper carried out a comparative assessment of the performance of LDA, and two of its variants, that is KLDA and MLDA in face recognition application. The face recognition system comprises of modules which involve face image acquisition, image preprocessing, feature extraction and feature classification for recognition. African face dataset from Ladoke Akintola University of Technology, Ogbomoso (LAUTECH) was used. The LDA techniques were used independently for feature extraction and the feature classification in all cases was achieved using Euclidean distance. The best among the three LDA techniques in face recognition was ascertained based on their performance. The rest of the paper is organized into the following: Section two presents review of relevant literatures to this research; Section three details the research methodologies employed in the comparative analysis; Section four present the results and Sections five summarized and concludes the paper.

2. LITERATURE REVIEW

2.1 Face Recognition

Face recognition is one of the most important applications of biometrics based authentication system in the last few decades. Face recognition is a type of recognition task pattern, where a face is categorized as either known or unknown after comparing it with the images of a known person stored in the database. Over the years, face recognition has found applications in security, criminal justice systems, image database investigation, surveillance, smart card applications, video indexing, human computer interaction, multimedia environment with adaptive human computer interface to mention, but a few.

Face recognition is a challenge, given the certain variability in information because of random variation across different people, including systematic variations from various factors such as lightening conditions, pose and so on [15]. The human face is an extremely complex and dynamic structure with characteristics that can significantly and quickly change in time. Face recognition involves a range of activities from various aspects of human life. Humans can recognize faces, but too many faces sometimes being hard to memorized, machine learning is now being improved to do this task. Researchers attempt to understand the architecture of the human face when building or developing face recognition systems. Atalay (1996) presented a face recognition system that heavily carries the characteristics of a typical pattern recognition system. The system was summarized in modules as follows [26]:

i. Acquisition module is the entry point of the face recognition process. It is the module in which the face image under consideration is presented to the system. An acquisition module can request a face image from several different environments such as well-illuminated environment.
ii. Pre-processing module by means of early vision techniques, face images are enhanced by using histogram equalization method and if desired, they are enhanced to improve the recognition performance of the system.
iii. Feature extraction module takes place after performing some pre-processing (if necessary), the normalized face image is presented to the feature extraction module in order to find the key features that are going to be used for classification.
iv. Classification module are used, with the help of a pattern classifier, extracted features of the face image is compared with the ones stored in a face library (or face database). After doing this comparison, face image is classified as either known or unknown. Training sets are used during the "learning phase" of the face recognition

process. The feature extractions and the classification modules adjust their parameters in order to achieve optimum recognition performance by making use of training sets. Face library or face database is a repository of face images which after some face images are being classified as "unknown", face images can be added to a library (or to a database) with their feature vectors for later comparisons. The classification module makes direct use of the face library [26].

2.2 Feature Extraction Techniques

Feature extraction is a very important field of image processing and face recognition. Fundamental component of characters is called features. The basic task of feature extraction and selection is to find out a group of the most effective features for classification; that is, compressing from high-dimensional feature space to low-dimensional feature space, so as to design classifier effectively [10]. Feature extraction process can be defined as the procedure of extracting relevant information from a face image. This information must be valuable to the later step of identifying the subject with an acceptable error rate. The feature extraction process must be efficient in terms of computing time and memory usage. The output should also be optimized for the classification step. Feature extraction involves several steps - dimensionality reduction, feature extraction and feature selection. These steps may overlap, and dimensionality reduction could be seen as a consequence of the feature extraction and selection algorithms. Both algorithms could also be defined as cases of dimensionality reduction [13].

Dimensionality reduction is an essential task in any pattern recognition system. The performance of a classifier depends on the amount of sample images, number of features and classifier complexity. One could think that the false positive ratio of a classifier does not increase as the number of features increases. However, added features may degrade the performance of a classification algorithm. This may happen when the number of training samples is small relatively to the number of features. This problem is called “curse of dimensionality” or “peaking phenomenon”.

A generally accepted method of avoiding this phenomenon is to use at least ten times as many training samples per class as the number of features. This requirement should be satisfied when building a classifier. The more complex the classifier, the larger should be the mentioned ratio [14]. This “curse” is one of the reasons why it’s important to keep the number of features as small as possible. The other main reason is the speed. The classifier will be faster and will use less memory. Moreover, a large set of features can result in a false positive when these features are redundant. Ultimately, the number of features must be carefully chosen. Too less or redundant features can lead to a loss of accuracy of the recognition system. There are two predominant approaches to the face recognition problem: geometric (feature based) and photometric (view based). As researcher interest in face recognition continued, many different algorithms were developed, such as Discrete Cosine Transform (DCT), Principal Components Analysis (PCA), Fisher Linear Discriminant Analysis (FLDA), and Elastic Bunch Graph Matching (EBGM).

2.3 Linear Discriminant Analysis (LDA)

Originally developed in 1936 by R.A. Fisher, discriminant analysis is a classic method of classification that has stood the test of time. Discriminant analysis often produces models whose accuracy approaches (and occasionally exceeds) more complex modern methods. Discriminant analysis can be used only for classification (that is with a categorical target variable), not for regression. The target variable may have two or more categories. It is also known as Fisher Discriminant Analysis (FDA) [21]. Dimensionality reduction is fundamentally important for analyzing high-dimensional data, and has received sufficient attention in the field of artificial intelligence [23]. The goal of dimensionality reduction is to embed the data into a low-dimensional subspace, while retaining the desired discriminant information. The pseudo code for LDA is depicted below:

The description of Fisher Linear Discriminant Analysis procedure is given below:

Given the data matrix $X = [x_1, x_2, \ldots, x_n), x_j \in \mathbb{R}^{d+1}$ with C classes, the purpose of LDA is to learn a linear transformation matrix $W \in \mathbb{R}^{d+m}$ $(m \ll d)$ to map the *d*-dimensional data x_j to a m-dimensional vector:

$$y_j = W^T x_j \tag{2.1}$$

FLDA supposes that an optimal transformation should push the data points from different classes far away from each other while pulling those within the same class close to each other. So the objective of FLDA can be written as

$$\max_W \frac{\sum_{i-1}^{C} n_i \left\|W^T(\mu^{'} - \mu)\right\|_2^2}{\sum_{i-1}^{C}\sum_{j=1}^{n_i} ||W^T(\mu^{'} - \mu)||_2^2} \tag{2.2}$$

where n_i is the number of samples in class *i*, μ^i is the mean of the samples in class *i*, μ is the mean of all the samples, and x_j^i is the *j*-th sample in class *i*. Denote the between-class scatter matrix S_b and the within-class scatter matrix S_w as in equation (2.3) and (2.4)

$$S_b = \sum_{i=1}^{C} n_i(\mu^i - \mu)(\mu^i - \mu)^T \tag{2.3}$$

$$S_w = \sum_{i=1}^{C}\sum_{j=1}^{n_i} (x_j^i - \mu^i)(x_j^i - \mu^i)^T \tag{2.4}$$

then the problem can be rewritten into a concise form:

$$\max_W \frac{tr(W^T S_b W)}{tr(W^T S_w W)} \tag{2.5}$$

Where t*r*() indicates the trace operator. Due to the complexity to solve the above trace ratio problem, many researchers transform it into a ratio trace form,

$$\max_W tr\left(\frac{W^T S_b W}{W^T S_w W}\right) \tag{2.6}$$

2.4 Kernel Linear Discriminant Analysis (KLDA)

KLDA is a generalization of Fisher Linear Discriminant Analysis (FLDA), a statistical method to find linear combinations of features (that is variables in a data set, or points in a trace) that characterize class separations. In particular, it outputs projection directions that maximize the ratio of between-group to within group scatter, so that ‘interesting’ variation may be concentrated into a reduced dimension space for further analysis. KLDA has been promoted as one of a number of methods to extract sensitive data dependent features from side-channel traces for some years [31]. However, because it only finds linear combinations, it is unable to locate the types of joint data dependencies exhibited by traces which have been protected by software masking. By contrast, the ‘kernel trick’ employed by KLDA allows to implicitly map the data into a higher dimensional feature space within which to perform the discriminant analysis, thereby extracting non-linear combinations of the sort that do yield sensitive information on further analysis [31]

.

Discriminant Analysis with Kernels LDA can be used to find optimal linear mappings of high dimensional data but is not applicable when the relevant information is known to be contained in non-linear combinations of points, as is the case for side-channel leakages of masked implementations. To extend FLDA to the non-linear case, we consider the problem in a feature space F induced by some mapping function (this mapping process is implicit as will be seen in the following subsection), $\Phi: R^n \rightarrow \mathcal{F}$. KLDA is used to find nonlinear directions by first mapping the data non-linearly by Φ into some feature space F within which to compute linear discriminants, thus implicitly yielding a non-linear discriminant in the input space [20]. To find such a discriminant, equation 2.7 is used:

$$J(\omega') = \frac{W^T S_B^{\Phi} \omega'}{W^T S_W^{\Phi} \omega'} \tag{2.7}$$

Where $\omega' \in \mathcal{F}$ and $S_B^{\Phi}$and S_W^{Φ} are the corresponding matrices in F.

$$S_B^{\Phi} = \sum_{m \epsilon M} n_m \left(\frac{1}{n_m} \sum_{m_i=m} P_i^{\Phi} - \frac{1}{N} \sum_{i=1}^{N} P_i^{\Phi}\right)^T \left(\frac{1}{n_m} \sum_{m_i=m} P_i^{\Phi} - \frac{1}{N} \sum_{i=1}^{N} P_i^{\Phi}\right) \tag{2.8}$$

$$S_W^{\Phi} = \sum_{m \epsilon M} \sum_{m_i=m} \left(P_i^{\Phi} - \frac{1}{n_m} \sum_{m_i=m} P_i^{\Phi}\right)^T \left(P_i^{\Phi} - \frac{1}{n_m} \sum_{m_i=m} P_i^{\Phi}\right) \tag{2.9}$$

where P_i^{Φ} is $\Phi(P_i)$ projection of P_i on *F* by Φ. For a properly chosen Φ an inner product <·,· >can be defined on *F*, which makes for a so-called ‘reproducing kernel Hilbert space’,

$$K(x, y) = < \Phi(x), \Phi(y) > \tag{2.10}$$

where K is known as the kernel function. Widely-used kernel functions include the Gaussian kernel $K(x, y) = \exp(-\|x - y\|^{2/c})$($\|\cdot\|$ is the 2-norm), and the polynomial kernel $K(x, y) = (x \cdot y)^{d'}$, for positive constants c and d' satisfying Mercer’s condition [25], as defined in [29].

2.5 Multiclass Linear Discriminant Analysis (MLDA)

If the number of classes is more than two, then a natural extension of Fisher Linear discriminant exists using multiple discriminant analysis [16]. As in two-class case, the projection is from high dimensional space to a low dimensional space and the transformation suggested still maximizes the ratio of intra-class scatter to the inter-class scatter. But unlike the two-class case, the maximization should be done among several competing classes. Suppose that now there are n classes. The intra-class matrix is calculated as:

$$\textstyle\sum_w = S_1 + \ldots + S_n = \sum_{i=1}^{n} \sum_{x \epsilon c_i} (x - \bar{x}_i)(x - \bar{x}_i)' \tag{2.11}$$

The inter-class scatter matrix slightly differs in computation and is given by $\sum_b = \sum_{i=1}^{n} m_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})'$ (2.12)

Where m_i is the number of training samples for each class, $\bar{x}_i$ is the mean for each class and$\bar{x}$ is total mean vector given by $\bar{x} = \frac{1}{m} \sum_{i=1}^{n} m_i \bar{x}_i$. After obtaining $\sum_b$. and $\sum_w$, the linear transformation Φ. It can be shown that the transformation Φ can be obtained by solving the generalized eigenvalue problem:

$$\textstyle\sum_b \Phi = \lambda \sum_w \Phi \tag{2.13}$$

It is easy to prove that the upper bounds of the rank of $\sum_b$. and $\sum_w$, are respectively m−n and n−1. Multiple discriminant analysis provides an elegant way for classification using discriminant features. If classification is required, instead of dimension reduction, there are a number of alternative techniques available. For instance, the classes may be partitioned, and a standard Fisher discriminant or LDA used to classify each partition. A common example of this is "one against the rest" where the points from one class are put in one group, and everything else in the other, and then LDA applied. This will result in C classifiers, whose results are combined. Another common method is pair-wise classification, where a new classifier is created for each pair of classes (giving C (C – 1)/2 classifiers in total), with the individual classifiers combined to produce a final classification.

2.6 Related Works

[8] proposed a technique that involved using FLDA for classification. The approach was termed Clustering based Discriminate Analysis (CDA) and achieved a recognition accuracy of 93% for three classes of expression. The conventional linear approach like LDA and PCA are straightforward and proficient on the grounds that they are linear. Notwithstanding, these are not appropriate for representing powerfully changing facial expressions in light of the fact that the changing expressions are characteristically non-linear.

[11] present a novel face recognition system that uses two-class linear discriminant analysis for classification. In this approach a single M-class linear discriminant classifier was divided into M two-class linear discriminant classifiers. This formulation provides many advantages like more discrimination between classes, simpler calculation of projection vectors and easier update of the database with new individuals. The proposed algorithm was tested on the CMU PIE and Yale face databases. Two-class LDA performs slightly better than the multi-class LDA, where there is only 2.22%, 10.29%, performance difference between the best classification scores of these two algorithms for Yale and CMU respectively. Significant performance improvements were observed, especially when the number of individuals to be classified increases.

[22] proposed an optimised fisher discriminant analysis for recognition of faces having black features. About 460 faces samples from 46 black African individuals (with and without tribal marks) were acquired. In the experiment, different sizes of gray scale images were used for recognition and performance accuracy of between 88 and 99% were obtained. Also, taken into consideration was the rate of identifying an image using the same number of images to test the face recognition system. The optimized fisher discriminant analysis was found to be efficient.

[7] proposed a face recognition system by Linear Discriminant Analysis (LDA). ORL face database consisting of ten different images each for 40 distinct subject is used for both training and testing. Three hundred and sixty images were used for training while forty images were used for testing. 37 of the images were correctly recognised while 3 were wrongly recognised to achieve an accuracy of 92.5%.

[2] investigated three PCA based face recognition system which involves PCA, PCA-ANN and BPCA (Binary PCA). They utilized 400 face images which is made up of four (4) facial expression images for 100 individuals. The experimental results revealed that PCA-ANN method achieved the best recognition accuracy of 94%.

[1] analysed the Performance of different Support Vector Machine kernels (Radial Basis Function, Linear Function, Quadratic Function and Polynomial Function) for face emotion recognition. A local African database of 714 face emotion images consisting of seven facial expression taken twice from 51 persons was used. The results obtained using the SVM multi-class classification scheme reveals that the Quadratic Function SVM kernel performs best for face emotion recognition with an average accuracy of 99.33%. However, despite the good performance achieved with higher dimensions the computation time is high.

[33] proposed a method which involved Distance Transform on a Kernel Discriminant Analysis DT_KDA to extraction, and the recognition using Kohonen SOM. The work involved two approaches. The first approach is a combination of KDA-DT-Kohonen, the

second is KDA-Kohonen and tested on two datasets: CALTECH and Computer Vision (CE1). The second dataset is used to describe the effect of rotation of the face and background. Extraction of facial features using KDA without DT was found to be more accurate as the Kohonen SOM network parameters for recognizing the face at CALTECH and CE1 dataset. The KDA-Kohonen techniques achieved 98.79% and 79.65 % using CE1 and CALTECH dataset respectively, while KDA-DT-Kohonen techniques achieved 92.78% and 76.09 % using CE1 and CALTECH dataset respectively.

In the above review LDA techniques had good performance in terms of the performance metrics used. Most of the work uses few parameters without requiring additional training or any parameter optimization. However, most of the existing techniques have issues within representing powerfully changing facial feature due to the fact that changing expressions of the face are characteristically non-linear. Also, there are issues with the computational efficiency with respect to training and testing times. Therefore, this research carried out a comparative analysis of some selected LDA techniques. The best among these techniques was determined based on the aforementioned performance metrics.

3. METHODOLOGY

3.1 Overview of the Methodology

In this paper work, three hundred and forty (340) static facial images were obtained using a digital camera. The acquired images were divided into training dataset and testing dataset. The acquired images were pre-processed after cropping and resizing them. Noise and other unwanted elements were removed from the images. The coloured images were converted into gray scales for time and memory management using function rgb2gray (RGB) in MATLAB Computing Toolbox. Normalization of the images was achieved through the application of histogram equalization techniques. The feature dimensionality reduction, separation and extraction of the pre-processed image was achieved by the application LDA techniques (LDA, KLDA and MLDA). Euclidian distance was used for similarity measurement between the tested images and the trained images. The results obtained was evaluated using recognition accuracy, precision, sensitivity, false positive rate and computation time to determine the performance of the techniques. Figure 1 depicts the scheme for evaluating the LDA techniques while Figure 2 depicts the flowchart of the procedure for training and testing face with LDA techniques.

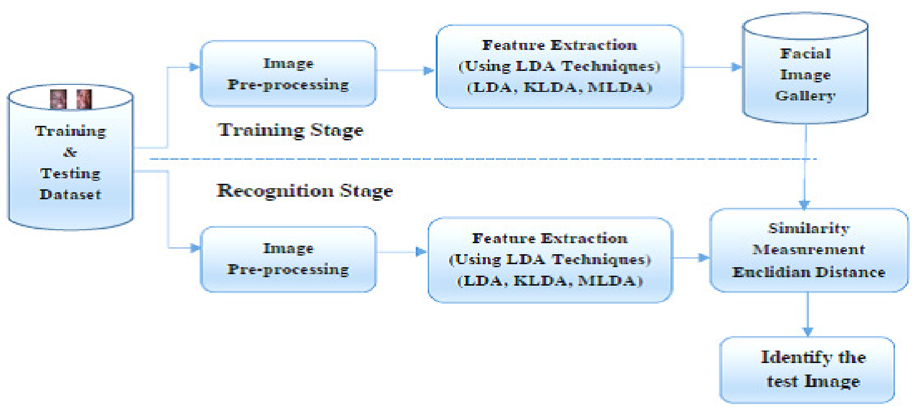

Figure 1: The Scheme for Evaluating the LDA Techniques

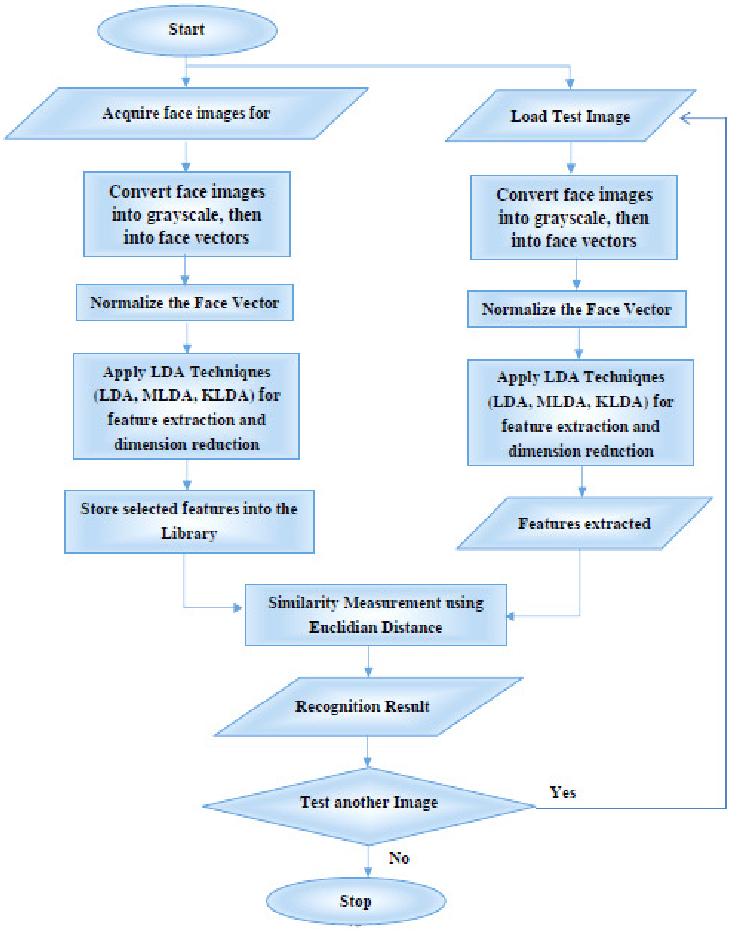

Figure 2: Flowchart for the Procedure of Training and Testing Face with LDA Techniques

3.2 Stages of the Face Recognition Scheme

a) ***Acquisition of Face Images:*** Three hundred and forty (340) images were taken with a digital camera at 1200 x 1200 pixel resolution. The original face images were downsized into a suitable pixel. Two hundred and forty (240) of the images were used for training while the remaining One Hundred (100) were used for testing.

b) ***Image Pre-processing:*** Image pre-processing has to do with actions such as image brightness, contrast alteration, image scaling, filtering, cropping and other operations that helps in the enhancement of images. In this phase, pre-processing was carried out by converting the coloured image into grayscale and normalizing of face vectors by computing the average face vector and deducting average face from each face vector. This was done to remove noise and other unwanted element from the face images. This stage helps to get rid of unwanted information that would have been extracted as features and reduces the work to be done during dimensionality reduction (feature extraction). Grayscale conversion is necessary to reduce the number of pixels.

c) ***Conversion of Face Images into Grayscale and Face Vector***: The image acquired from the digital camera was coloured images in three-dimensional form (3-D). The coloured images were converted into grayscale using the MATLAB function rgb2gray so as to reduce processing time being a two-dimensional matrix. Each of the grayscale images were expressed and stored in form of matrix in MATLAB which was converted to vector image for further processes. The conversion to face vector was made to aid the normalization process.

d) ***Normalization of Face Image***: The normalization of the images was carried out by applying histogram equalization technique to the converted grayscale images to improve the contrast in the images by stretching out the intensity range. This enhances the brightness in the grayscale images for clearer view of the face of each subject. Normalization phase removes any common features that all the face images shared together, so that each face images is left with unique features. The common features were discovered by finding the average face vector of the whole training set (face images). Then, the average face vector was subtracted from each of the face vectors which results into a normalized face vector.

e) ***Feature Extraction***: Significant collection of basic parameters (face features) that best illustrate the specific array of face images was extracted from the pre-processed image of each subset and was used to discriminate between them. The extracted face features was encoded and stored as weight vectors for each face images in order to compare it to other images in the training dataset. Three variants of Fisher Linear Discriminant Analysis techniques (i.e. LDA, KLDA and MLDA) were employed independently in this study to extract features and reduce the dimension sizes of images. The resultant feature representation extracted by these techniques presented a suitable platform to identify a test image. LDA produces an optimal linear discriminant function which maps the input into the classification space in which the class identification of this sample is decided based on some metric such as Euclidean distance. Thus the objective of LDA is to find the optimal projection, so that the ratio of determinants of between-class and the within class scatter matrices of the projected samples reaches its maximum. Linear Discriminant Analysis projects into a subspace that maximizes the between class scatter while minimizing within class scatter of the projected data. LDA improves the generalization capability by decomposing into a simultaneous diagonalization of the two within- class covariance matrices. The robustness of the LDA procedure depends on whether the within-class scatter captures reliable variations for a specific class or not.

3.3 Euclidean Distance

The extracted features by the LDA techniques i.e. LDA, KLDA and MLDA were classified using Euclidean Distance. It was employed to measure the similarity between the test vector and the reference vectors in the gallery. Euclidean distance is defined as the straight-line distance between two points. For *N*-dimensional space, the Euclidean distance between two any points' *pi* and *qi* is given by equation (3.1):

$$D(x,y) = \sqrt{\sum_{i=1}^{N}(x_i - y_i)^2} \tag{3.1}$$

Where $x_i$and y_i is the coordinate of *x and y* in dimension *i*.

3.4 Evaluation Measures

The performance of the variants of LDA techniques on both trained and recognized faces was evaluated based on recognition accuracy, false positive rate, sensitivity, specificity and average recognition time. Confusion matrix was used to determine the value of the performance metrics. It contains “True Positive (TP), False Positive (FP), False Negative (FN) and True Negative (TN).” TP contains amount of entries for the tuple that correctly identified as positive. FP contains the amount entries for the tuples which are negative but predicted as positive. TN is the number of tuples that are negative and predicted as negative. FN is the number of tuples that are positive but predicted as negative. Sensitivity, specificity and accuracy will be calculated using these terms.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3.2}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{3.3}$$

$$\text{False Positive Rate} = \frac{\text{FP}}{\text{TN} + \text{FP}} = 1 - Specificity \tag{3.4}$$

$$\text{Overall Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{3.5}$$

$$\text{Average recognition time} = \frac{\text{Total Recognition Time}}{\text{Number of recognized faces}} \tag{3.6}$$

The graphical representation of the relationship between the dimension size and the average training time as well as that of threshold values and the recognition time was plotted by MS-excel (2016). The regression analysis base on the computation time against the dimension size and the threshold values was also conducted using MS-excel (2016). Furthermore, the IBM SPSS Statistic version 21 was used to conduct the statistical analysis.

3.4 Implementation in MATLAB

The applied techniques were implemented using MATLAB R2015a version on Windows 10 Enterprise 64-bit operating system, Intel®Pentium® CPU T4500@2.30GHZ Central Processing Unit, 4GB RAM and 500 Gigabytes hard disk drive. An interactive Graphic User Interface (GUI) was developed with a real time database consisting of 340 face images. The techniques will be evaluated based on the aforementioned performance metrics. The model was experimented by taken into consideration the face recognition in 50 by 50, 100 by 100, 150 by 150 and 200 by 200-pixel resolution.

4. RESULTS AND DISCUSSION

4.1 Summary of results

A couple of screenshots of the GUI of the implementation environment (MATLAB) is depicted Figures 3 and 4. The time spent by each LDA technique for training the dataset is shown in Table 1(a), Table 1(b) and Table 1(c). The time spent increases as the dimension size of the images increases, which implies that the time consumed depends on the features in the training set for LDA, KLDA and MLDA. The average training time generated by application of LDA after two trial for images at 50 by 50 pixel resolution is 469.16 s, 100 by 100 pixel resolution is 591.42 s, 150 by 150 pixel resolution is 908.92 s, 200 by 200 pixel resolution is 1311.76 s as presented in Table 1(a). Similarly, the average training time generated by application of KLDA for image of at at 50 by 50 pixel resolution is 488.46 s, 100 by 100 pixel resolution is 618.05 s, 150 by 150 pixel resolution is 977.15 s, 200 by 200 pixel resolution is 1393.24 s as presented in Table 1(b). Also, the average training time generated by application of MLDA for image of at at 50 by 50 pixel resolution is 431.47 s, 100 by 100 pixel resolution is 550.97 s, 150 by 150 pixel resolution is 855.12 s, 200 by 200 pixel resolution is 1191.55 s as presented in Table 1(c). The result shows that the MLDA among other is less computationally expensive in terms of training time compared to the LDA and KLDA model.

Figure 3: MATLAB GUI Showing Results of the Training Stage of Face Recognition

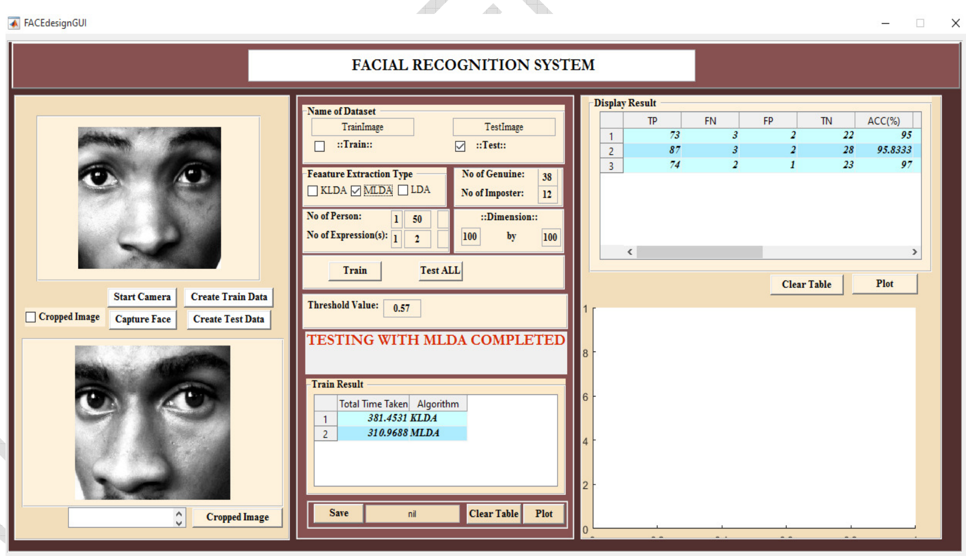

Figure 4: MATLAB GUI Showing Results of Testing Stage of Face Recognition

4.2 Experimental results

The LDA, KLDA and MLDA model were experimented by implementing the facial expression recognition using 200 x 200-pixel resolution. The system was tested and evaluated using the following performance metric: sensitivity, specificity, false positive rate, recognition accuracy and computation time. All performance metrics were analysed using by using a square dimension pixel resolution stated above at different threshold values.

Table 1: Average Training Time at Different Resolutions for LDA, KLDA and MLDA

(a) With LDA

Dimension Size	Time1(s)	Time2(s)	Average Time (seconds)
50 by 50	462.67	475.64	**469.16**
100 by 100	587.56	595.27	**591.42**
150 by 150	902.89	914.94	**908.92**
200 by 200	1318.22	1305.29	**1311.76**

(b) With KLDA

Dimension Size	Time1(s)	Time2(s)	Average Time (seconds)
50 by 50	496.26	480.65	**488.46**
100 by 100	625.41	610.69	**618.05**
150 by 150	970.95	983.34	**977.15**
200 by 200	1390.49	1395.99	**1393.24**

(c) With MLDA

Dimension Size	Time1(s)	Time2(s)	Average Time (seconds)
50 by 50	427.39	435.54	**431.47**
100 by 100	558.14	543.79	**550.97**
150 by 150	860.67	849.57	**855.12**
200 by 200	1193.28	1189.81	**1191.55**

Table 2: Experimental Results for MLDA, KLDA and LDA

(a) MLDA at 200 x 200-pixel resolution

Threshold	FPR (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Recognition Time (sec)
0.25	20.00	98.57	80.00	93.00	57.56
0.35	13.33	98.57	86.67	95.00	58.89
0.46	6.67	97.14	93.33	96.00	59.01
0.57	3.33	97.14	96.67	97.00	58.65

(b) KLDA at 200 x 200-pixel resolution

Threshold	FPR (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Recognition Time (sec)
0.25	20.00	97.14	80.00	92.00	64.23
0.35	13.33	97.14	86.67	94.00	64.89
0.46	10.00	95.71	90.00	94.00	63.89
0.57	6.67	95.71	93.33	95.00	63.67

(c) LDA at 200 x 200-pixel resolution

Threshold	FPR (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Recognition Time (sec)
0.25	26.67	95.71	73.33	89.00	67.89
0.35	16.67	94.29	83.33	91.00	68.45
0.46	10.00	92.86	90.00	92.00	68.02
0.57	6.67	92.86	93.33	93.00	67.98

4.2.1 Experimental Results for MLDA

Table 2(a) presented the result obtained by the MLDA at 200 x 200-pixel resolution at threshold value of 0.25, 0.35, 0.46 and 0.57 with respect to the performance metrics. The table reveals that the performance of MLDA varies with change in the threshold value. Also, it was discovered that accuracy, specificity increases with increase in threshold value while the false positive rate and sensitivity decreases with increase in the threshold value. However, the optimum performance was achieved at threshold value of 0.57. The MLDA achieved a false positive rate of 3.33%, sensitivity of 97.14%, specificity of 96.67% and accuracy of 97.0% at 58.65 seconds. The table also shows that the computation time is within the range of 57.56 to 59.65 seconds with increase in the threshold values.

4.2.2 Experimental results for KLDA

Table 2(b) presented the result obtained by the KLDA at 200 x 200-pixel resolution at threshold value of 0.25, 0.35, 0.46 and 0.57 with respect to the performance metrics. The table reveals that the performance of KLDA varies with change in the threshold value. Also, it was discovered that accuracy, specificity increases with increase in threshold value while the false positive rate and sensitivity decreases with increase in the threshold value. However, the optimum performance was achieved at threshold value of 0.57. The KLDA achieved a false positive rate of 6.67%, sensitivity of 95.71%, specificity of 93.33% and accuracy of 95.0% at 63.67 seconds. The table also shows that the computation time is within the range of 63.63 to 64.89 seconds with increase in the threshold values.

4.2.3 Experimental results for LDA

Table 2(c) presented the result obtained by the LDA at 200 x 200-pixel resolution at threshold value of 0.25, 0.35, 0.46 and 0.57 with respect to the performance metrics. The table reveals that the performance of LDA varies with change in the threshold value. Also, it was discovered that accuracy, specificity increases with increase in threshold value while the false positive rate and sensitivity decreases with increase in the threshold value. However, the optimum performance was achieved at threshold value of 0.57. The LDA achieved a false positive rate of 6.67%, sensitivity of 92.86%, specificity of 93.33% and accuracy of

93.0% at 67.98 seconds. The table also shows that the computation time is within the range of 67.89 to 68.45 seconds with increase in the threshold values.

4.2.4 Comparison Results between MLDA, KLDA and LDA

Table 3 shows a combined result of MLDA, KLDA and LDA at the threshold value of 0.57 with respect to all metrics at 200 by 200-pixel resolution. All result obtained in Table 3 presume that MLDA model has a lower recognition time compared with the corresponding KLDA and LDA model irrespective of threshold value.

Similarly, Recognition accuracy, sensitivity, false positive rate and specificity of MLDA, KLDA and LDA model are compared at 200 by 200-dimensional size; the study discovered that MLDA model has better performance in accuracy, specificity and false positive rate than KLDA and LDA model as enumerated in Table 3. The recognition accuracy of 97.0% with MLDA, 95.0% with KLDA and 93.0 % with LDA model. The MLDA model have a specificity of 96.67%, false positive rate of 3.33% and sensitivity of 97.14% at 58.65; the KLDA model have a specificity of 93.33%, false positive rate of 6.67% and sensitivity of 95.71% at 63.67 while the LDA model have a specificity of 93.33%, false positive rate of 6.67% and sensitivity of 92.86% at 67.98. Hence, MLDA outperformed KLDA and LDA.

Table 3: MLDA, KLDA and LDA at 200 x 200-pixel Resolution and 0.57 Threshold Value

Algorithm	FPR (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Recognition Time (sec)
MLDA	3.33	97.14	96.67	97.00	58.65
KLDA	6.67	95.71	93.33	95.00	63.67
LDA	6.67	92.86	93.33	93.00	67.98

4.3 Discussion of Results

The experimental results discussion in terms of training and recognition computation time analysis, evaluation of other performance metrics and statistical analysis is presented in this section.

4.3.1 Computation Time Analysis

The results shown in Table 1 shows that the MLDA model trains the dataset much faster than the KLDA and LDA model. Therefore, the MLDA is less computationally expensive compared to both KLDA and the LDA model. The training time increases with increase in the features of the training set. Figure 5 shows the graph of average training time against the dimension size. The relationship between the average training time (T_t) and the dimension size (dm) is found to be linear with a high correlation coefficient for MLDA, KLDA and LDA model as shown in equations 4.1, 4.2 and 4.3 respectively.

$$T_t = 0.0207dm + 369.02 \qquad R^2 = 0.9967 \tag{4.1}$$

$$T_t = 0.0247dm + 405.84 \qquad R^2 = 0.9962 \tag{4.2}$$

$$T_t = 0.0229dm + 390.44 \qquad R^2 = 0.997 \tag{4.3}$$

Similarly, Figure 6 shows the graphs which depict the relationship between the average recognition time and the threshold values for MLDA, KLDA and LDA model respectively. From the graph; the relationship between the recognition time (T_R) and the threshold values (*th*) is found to be quadratic with a high correlation coefficient for MLDA and polynomial of the third order with a high correlation coefficient for both KLDA and LDA model as shown in equation 4.4, 4.5 and 4.6 respectively.

$$T_R = -29.849th^2 + 64.566th + 53.18 \qquad R^2 = 0.9915 \qquad (4.4)$$

$$T_R = 260.75th^3 - 327.61th^2 + 127.17th + 49.332 \quad R^2 = 0.9999 \qquad (4.5)$$

$$T_R = 143.86th^3 - 182.42th^2 + 72.621th + 59.21 \quad R^2 = 0.9999 \qquad (4.6)$$

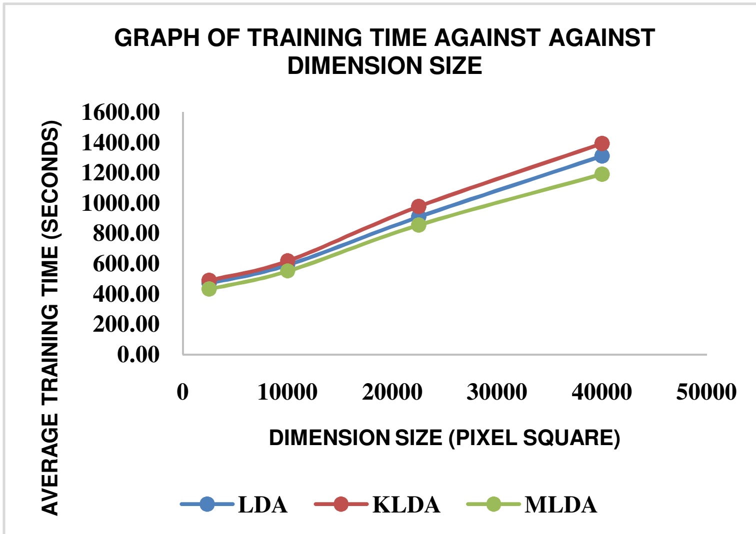

Figure 5: Relationship between Average Training Time (seconds) and Dimension size (Pixel Square)

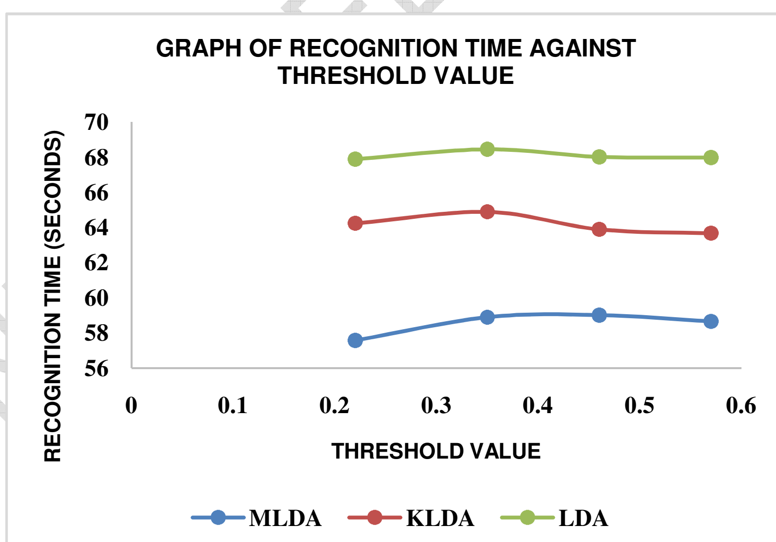

Figure 6: Relationship between Recognition Time and Threshold Values

With the computation time analysis, it was discovered that MLDA is less computationally expensive in terms of training and time recognition time compared to KLDA and LDA. The KLDA used more time to train the dataset.

4.6.2 Discussion Based on Performance Metrics

The results obtainable in Table 2 show the performance of MLDA, KLDA and LDA model. The results show that there is significant variation in the performance metrics with increase in threshold value and the best result is obtained at the threshold value of 0.57 across all metrics (false positive rate, specificity, sensitivity and accuracy) for MLDA, KLDA and LDA. Therefore, the performance of these techniques is dependent on the threshold value. It can be inferred from the results based on the performance metrics that the MLDA model gave an increased 2.0% recognition accuracy, 3.34% specificity, 1.43% sensitivity and a decreased FPR of 3.34% over the KLDA model at 0.57 threshold value. Similarly, MLDA model gave an increased 4.0% recognition accuracy, 3.34% specificity, 4.28% sensitivity and a decreased FPR of 3.34% over the LDA model at 0.57 threshold value. Hence, MLDA outperformed KLDA and LDA in terms of FPR, recognition accuracy, specificity and sensitivity.

The result achieved in this study is in line with the work of [24] which states that the variation in each of the variant of linear discriminant-based algorithms will have a varying performance in face recognition application due to improvement on the basic LDA. The results reveal that both KLDA and MLDA outperformed the basic LDA with MLDA having the optimum performance. Hence, the improvement on basic LDA improves the performance in facial recognition system. Nevertheless, the work of [11] proved otherwise. They reported that other two classes of LDA outperformed the multi-class LDA.

In view of the results, the MLDA is more accurate, specific and sensitive with minimal false positive than KLDA and LDA. Therefore, MLDA gave an improved accuracy, Sensitivity, specificity and false positive rate than KLDA and LDA.

4.6.3 Statistical Analysis of Facial Recognition Rates

Statistical analysis was conducted on the result obtained in this study. Accuracy and sensitivity were considered for analysis. The result in Table 2 shows that the MLDA has a higher recognition rate than the corresponding KDLA and LDA. A t-test values was measured between the accuracy of MLDA and KLDA as well as MLDA and LDA. The paired t-test analysis conducted between accuracy of MLDA and KLDA reveals a small mean difference ($\mu = 1.50$). Nevertheless, the result confirmed that the MLDA is statistically significant at $< 0.05; P = 0.014 \;\; and\; t\; value = 5.196$. Also, a t-test values was measured between the accuracy of MLDA and LDA. The paired t-test analysis conducted between MLDA and LDA reveals a small mean difference ($\mu = 3.75$). Nevertheless, the result confirmed that the MLDA is statistically significant at $< 0.01; P = 0.001 \;\; and\; t\; value = 15.0$. The t-test result further validates the fact the MLDA outperformed both KLDA and LDA in terms of recognition accuracy. Furthermore, a t-test values was measured between the sensitivity of MLDA and KLDA as well as MLDA and LDA. The paired t-test analysis conducted between MLDA and KLDA reveals a small mean difference ($\mu = 1.93$). Nevertheless, the result confirmed that the MLDA is statistically significant at $< 0.01; P = 0.007 \;\; and\; t\; value = 6.686$. Also, a t-test values was measured between the sensitivity of MLDA and LDA. The paired t-test analysis conducted between MLDA and LDA reveals a small mean difference ($\mu = 3.925$). Nevertheless, the result confirmed that the MLDA is statistically significant at $< 0.01; P = 0.002 \;\; and\; t\; value = 11.056$. The t-test result further validates the fact the MLDA outperformed both KLDA and LDA in terms of sensitivity.

5. CONCLUSION

This paper evaluated the essential features of variant of LDA face recognition system. Two hundred and forty (240) facial images were trained and One Hundred (100) images were used to test each of the LDA techniques model at different threshold value. The experimental results obtained revealed that MLDA outperformed the KLDA and LDA in terms of recognition accuracies, specificity, FPR, training and recognition computation time. In view of this, a face recognition system based on MLDA would produce a more reliable security surveillance system than KLDA and LDA. It should be considered in building a truly robust face recognition system where high recognition accuracy and computational efficiency must not be compromised. Future work can be carried out by investigating the performance of each of variant of LDA on a classifier such as Support Vector Machine (SVM), Artificial Neural network (ANN), Hidden Markov Model (HMM) and others. Furthermore, the performance of Hybrid of MLDA and a suitable evolutionary search algorithm like Ant Colony Optimization (ACO), Evolutionary Programming (EP), Genetic Programming (GP), Differential Evolution (DE) and Artificial Immune Systems (AIS) can be considered as subject for future research.

6. REFERENCES

[1] Adeyanju, I. A., Omidiora E. O. and Oyedokun O. F. (2015). Performance Evaluation of Different Support Vector Machine Kernels for Face Emotion Recognition. *In SAI Intelligent Systems Conference (IntelliSys) IEEE,* pp. 804-806.

[2] Aluko, J. O., Omidiora E. O., Adetunji A. B. and Odeniyi O. A. (2015). Performance Evaluation of Selected Principal Component Analysis-Based Techniques for Face Image Recognition. *International Journal of Scientific & Technology Research*, 4 (1), pp. 1-7.

[3] Atalay, I., (1996). Face Recognition Using Eigenfaces. Istanbul:Istanbul Technical University, pp. 1-20.

[4] Babatunde R. S., Olabiyisi S. O., Omidiora E. O. and Ganiyu R. A. (2015). Local Binary Pattern and Ant Colony Optimization Based Feature Dimensionality Reduction Technique for Face Recognition Systems. *British Journal of Mathematics & Computer Science,* 11(5), pp. 1-11.

[5] Bakshi U. and Singhal R. (2014). A survey on face detection methods and feature extraction techniques of face recognition. *International Journal of Emerging Trends and Technology in computer science (IJETTCS)*. 3(3): 233-237.

[6] Belhumeur P. N., Hespanha J. P. and Kriegman D. J. (1997). Eigenfaces vs.Fisherfaces: Recognition using class specific linear projection. *IEEETrans. Pattern Anal. Mach. Intell.*, 19(7): pp. 711–720.

[7] Bhattacharyya S. K. and Rahul K. (2013). Face Recognition by Linear Discriminant Analysis. *International Journal of Communication Network Security, ISSN: 2231 – 1882*, 2(2); pp. 31-35.

[8] Chen, X. and Huang T. (2003). Facial Expression Recognition: A Clustering-Based Approach. *Pattern Recognition Letters*, 24, pp. 1295-1302.

[9] Chin T. J., Schindler K., and Suter D. (2006). Incremental kernelsvd for face recognition with image sets. In *IEEE Automatic Face and Gesture Recognition, 7th International Conference*, pp 461–466.

[10] Dong-ping T. (2013). A Review on Image Feature Extraction and Representation Techniques. *International Journal of Multimedia and Ubiquitous Engineering*. 8(4): pp 385-396.

[11] Ekenel H. K. and Stiefelhagen R., "Two-class Linear Discriminant Analysis for Face Recognition", IEEE Signal Processing and Communications Applications Conference, Eskişehir, Turkey, June 2007.

[12] Etemad K. and Chellappa R. (1997). Discriminant analysis for recognition of human face images.*J. Opt. Soc. Amer. A*, 14(1): pp. 1724–1733.

[13] Ion. M, (2010) “Face recognition algorithms” A handbook of biometrics springer.pg1-22 ISBN. 978-0-387-71040-2.

[14] Jain A. K., Duin R. P. W. and Mao J. (2000). Statistical Pattern Recognition: A Review. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(1), pp. 4-37.

[15] Jain, A. K. and Li S. Z., 2011. Handbook of face recognition. New York: springer.

[16] Johnson R.A. and Wichern D. W. (1988) Applied Multivariate Statistical Analysis. 2nd Edition, John Wiley & Sons Inc., New York.

[17] Liang X. and Lin (2016). Maximal Margin Local Preserving Median Fisher Discriminant Analysis for Face Recognition. *Journal of Software.* 11(12): pp. 1172-1180.

[18] Liu C. and Wechsler H., (2002). Gabor feature based classification using theenhanced fisher linear discriminant model for face recognition. *IEEETrans. Image Process.*, 11(4): pp. 467–476.

[19] Medjahed S. A. (2015). A Comparative Study of Feature Extraction Methods in Images Classification. *I.J. Image, Graphics and Signal Processing.* 3: 16-23

[20] Mika S., Ratsch G., Weston J., Scholkopf B. and Mullers, K. R. (1999). Fisher Discriminant Analysis with Kernels. *In: Neural Networks for Signal Processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop*. pp. 41–48.

[21] Ohol R. and Ohol S., (2017). Linear Discriminant Analysis for Human Face Recognition. *International Research Journal of Engineering and Technology (IRJET).* 4(8):pp. 1-3.

[22] Omidiora E. O., Fakolujo A. O., Ayeni R. O. and Adeyanju I. A. (2008). Optimised Fisher Discriminant Analysis for Recognition of Faces Having Black Features. *Journal of Engineering and Applied Sciences*, 3 (7), pp. 524-531.

[23] Peng X., Lu J., Yi Z., and Rui Y (2016). Automatic Subspace Learning via Principal Coefficients Embedding. *IEEE Trans. Cybern.*, 1(99): pp. 1–14.

[24] Prince S. J. and Elder, J. H. (2007). Probabilistic Linear Discriminant Analysis for Inferences about Identity. *IEEE 11th international conference on Computer Vision* 1-8.

[25] Saitoh, S. and Sawano, Y. (2016). Theory of Reproducing Kernels and Applications, vol. 44. Springer 2016.

[26] Samiksha A., Pallavi K., and Shashikant G. (2014). Facial Expression Recognition Techniques: A Survey; in: *International Conference on Electrical, Electronics, Computer Science and Mathematics Physical Education & Management, (ICEECMPE).* pp. 2-5.

[27] Sarode N. S. and Patil A. M. (2015). Iris Recognition using LBP with Classifiers-KNN and NB*. International Journal of Science and Research (IJSR). 4(1):* pp. 1905-1909.

[28] Savithiri G. and Murugan A. (2011). Performance Analysis on Half Iris Feature Extraction using GW, LBP and HOG international Journal of Computer Applications (22)2, pp 27-32

[29] Schölkopf B., Smola A. and Müller K. R. (1998). Nonlinear Component Analysis as A KernelEigenvalue Problem. *Neural computation* 10(5): pp. 1299–1319.

[30] Shruti B. (2014). Feature Extraction of Face Using Various Techniques. *Department of Computer Science and Engineering National Institute of Technology Rourkela,* pp.26-27.

[31] Standaert F.X. and Archambeau C. (2008). Using subspace-based template attacks to compareand combine power and electromagnetic information leakages. *In: International Workshop on Cryptographic Hardware and Embedded Systems.* pp. 411– 425.

[32] Standaert F.X., Gierlichs B. and Verbauwhede I. (2008): Partition vs. comparison sidechannel distinguishers: An empirical evaluation of statistical tests for univariate side-channel attacks against two unprotected CMOS devices. *In: International Conference on Information Security and Cryptology.* pp. 253–267.

[33] Wijaya B. A., Husein A. M., Harahap M. and Harahap M. K. (2017) Implementation Distance Transform Method in Kernel Discriminant Analysis for Face Recognition Using Kohonen SOM, International Journal of Engineering Research & Technology, 6(10), pp 28 -31.

[34] Yu H. and Yang J. (2001). A Direct LDA Algorithm for High Dimensional data-with Application to Face Recognition. *Pattern recognition*, 34(10):2067-2070.