

Research Article

Facial Expression Recognition Based on LDA Feature Space Optimization

Fanchen Zheng 

College of Science, China Agricultural University, Beijing 100083, China

Correspondence should be addressed to Fanchen Zheng; 2020310020116@cau.edu.cn

Received 4 May 2022; Revised 3 July 2022; Accepted 4 July 2022; Published 29 August 2022

Academic Editor: Shahid Mumtaz

Copyright © 2022 Fanchen Zheng. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the development of artificial intelligence, facial expression recognition has become an important part of the current research due to its wide application potential. However, the qualities of the face features will directly affect the accuracy of the model. Based on the KDEF face public dataset, the author conducts a comprehensive analysis of the effect of linear discriminant analysis (LDA) dimensionality reduction on facial expression recognition. First, the features of face images are extracted respectively by manual method and deep learning method, which constitute 35-dimensional artificial features, 128-dimensional deep features, and the hybrid features. Second, LDA is used to reduce the dimensionality of the three feature sets. Then, machine learning models, such as Naive Bayes and decision tree, are used to analyze the results of facial expression recognition before and after LDA feature dimensionality reduction. Finally, the effects of several classical feature reduction methods on the effectiveness of facial expression recognition are evaluated. The results show that after the LDA feature dimensionality reduction being used, the facial expression recognition based on these three feature sets is improved to a certain extent, which indicates the good effect of LDA in reducing feature redundancy.

1. Introduction

With the development of the information society, as one of the subjects for human information exchange, the accurate recognition of facial expression will help to improve the efficiency of people's information exchange and to save time and cost. As one of the most primitive ways of information interaction, facial expression has been preserved to this day and plays an irreplaceable role. People's facial expressions can reveal their attitudes to the received information, the emotions toward people or things, and even their inner thoughts. Therefore, the technology of facial expression recognition can be applied to various scenarios, including human-computer interaction, as well as the security and medical fields, such as public security control, fraud behavior analysis, and patient emotion analysis. The method of facial expression recognition has become a research hotspot in the field of computer vision.

Facial expression recognition can be roughly described as analyzing the data of face images by computer, extracting the useful feature information for recognition, and using this information to determine the category of the images. However, as the facial features are extracted and constructed by a variety of feature extraction methods, they are often of high dimensions. If the high-dimensional features are directly used in the subsequent recognition tasks, then it will bring great difficulties for the recognition task. The reason is that although the high-dimensional features contain the complete face information, they have a large number of redundant features. As the number of samples increases, the redundant information will greatly increase the computational resource overhead but reduce the performance of the recognition model. When the amount of data cannot support the feature dimension, a "dimension disaster" will occur. The performance of the recognition model drops greatly or even becomes ineffective. Therefore, it is a key step to reduce the dimension of high-dimensional features in face

recognition technology, which has extremely important practical significance.

The accurate recognition of facial expression recognition is inseparable from the extraction of multidimensional features, so redundant features are inevitable. Although the LDA feature dimensionality reduction method has been widely used in the field of face recognition, the previous work had not analyzed the dimensionality reduction effect of different types of feature sets on expressions. Based on the KDEF face public dataset, this study constructs three types of feature sets such as face traditional artificial feature sets, deep learning feature sets, and hybrid feature sets. Furthermore, it performs LDA feature dimension reduction respectively and uses machine learning models to classify the facial expressions. The experimental results show that dimensionality reduction of hybrid features can achieve the best accuracy of facial expression classification, which shows the importance of multidimensional hybrid feature extraction for machine learning tasks.

In this study, it introduces the positioning of human facial key points, the definition of action units, and the construction of the coding system in Section 2. In Section 3, the process of image correction, feature extraction, feature set construction, and feature dimensionality reduction are introduced in detail. In Section 4, the classification effects of different feature sets under different classification methods and dimensionality reduction methods are compared and analyzed. The conclusions and future studies are presented in Section 5.

2. Related Work

The key of facial expression recognition is to construct the concise and accurate facial feature descriptors. The method based on the complete face recognition generally uses the traditional feature descriptors, such as Gabor [1], local binary pattern (LBP) [2], and histogram of oriented gradient (HOG) [3], to extract and construct the global feature set of face images. However, in the actual scenes, the face information in the collected images is often incomplete due to various factors, such as the information loss caused by the shooting angle and the face occlusion caused by glasses, masks, or beards. As the method cannot extract enough face information for recognition, it is difficult to improve the recognition accuracy. In this regard, researchers propose a method to extract high-dimensional face features by the key points of the face [4]. The core of this method is to accurately locate the key points of the face and correct them. The research on the facial key points can be traced back to the facial action coding system (FACS) [5]. FACS is developed by Ekman and Friesen to study the correspondence between facial muscle movements and different expressions. It is a system for classifying human facial movements through facial appearance. The movement of the facial muscles is encoded by FACS in accordance with the subtle changes of the face. It is deconstructed into specific action units (AUs) that generate expressions, which are independent but interconnected. The codes of 46 main action units and their definitions are listed in Table 1.

The features of these facial AUs and the main regions they control can not only be used to distinguish faces but also to reflect the related facial expressions [6]. Since then, more studies have begun to optimize the key feature descriptors of faces. Belhumeur et al. [7] used the global model to generate local position information as latent variables, derived a Bayesian objective function to optimize these local positions, and used the literature research method [8] to align the faces. Literature [9] introduced a method for constructing facial feature descriptors based on orthogonal difference-local binary pattern (OD-LBP). First, 3 gray level differences are calculated for each orthogonal location by subtracting the 2 nearest neighbors and the center pixel from their respective orthogonal values. Second, the respective differences of gray levels are counted. Finally, a vector is generated by concatenating the binary patterns produced by the two orthogonal groups. Kazemi and Sullivan [10] proposed a general gradient-boosted framework to learn the set of regression trees. This framework used the set of regression trees to estimate the location of face descriptors directly from a sparse subset of pixel intensities and then extracted the face features according to the location.

In the process of facial expression recognition, in order to extract more comprehensive face features, the feature dimension is often too high because of the high complexity of facial image data. Therefore, it is usually necessary to reduce the dimension of the extracted features first to reduce the feature redundancy. Principal component analysis (PCA) [11] and linear discriminant analysis (LDA) [12] are widely used because of their simplicity and convenience. Principal component analysis (PCA) is a method based on the Karhunen-Loeve (K-L) transform of statistical theory [13]. Mondal and Bag [14] used PCA to reduce the dimension of the extracted features and combined with the minimum distance classification system to realize face recognition. Zhuang and Guan [15] combined Gabor wavelet and PCA for feature extraction, so as to deal with the full illumination of face images. Linear discriminant analysis (LDA) obtains the optimal projection direction by finding the extreme value of the Fisher criterion function, so that after the sample is projected in this direction, the interclass dispersion is the largest and the intraclass dispersion is the smallest. Fisher's criterion function was first proposed by Fisher in 1936. It was originally used to solve the problem of binary classification and then extended to multi-classification. Bodini et al. [16] combined the effectiveness of deep convolution neural network (DCNN) feature with the feature dimensionality reduction capability of LDA to solve the problem of face recognition for a single person and single sample. Benouareth [17] proposed a facial feature extraction method that combines likelihood sufficient dimension reduction (LSDR) and LDA to handle the face recognition under pose and illumination changes.

Although the LDA feature dimensionality reduction method has been widely used in the field of face recognition, the impact of LDA feature dimensionality reduction on the face recognition under different feature sets has not been evaluated. In this regard, according to the KDEF face public data set, this study conducts a comprehensive analysis for

TABLE 1: Codes and definitions of main action units.

| Action unit | Definition | Action unit | Definition | Action unit | Definition |
|-------------|-----------------------------------|-------------|---|-------------|---------------------------------------|
| AU01 | Raise the inner eyebrows | AU13 | Pull the corners of the mouth up | AU25 | Separate the lips to expose the teeth |
| AU02 | Raise the outer eyebrows | AU14 | Shrink the corners of the mouth towards the teeth | AU26 | Separate the lips to see the tongue |
| AU04 | Droop the eyebrows overall | AU15 | Pull the corners of the mouth down vertically | AU27 | Separate the lips to see the throat |
| AU05 | Raise the upper eyelids | AU16 | Pull the lower lip down | AU28 | Suck the lips to cover the teeth |
| AU06 | Lift up the cheeks | AU17 | Squeeze the lower lip upward | AU41 | Droop the upper eyelid slightly |
| AU07 | Constrict the eyes | AU18 | Wrinkle the mouth in the middle | AU42 | Droop the upper eyelid |
| AU09 | Shrink and lift the nose | AU20 | Pull the lips backwards | AU43 | Close the eyes |
| AU10 | Raise the upper lip | AU22 | Purse the lips into a funnel | AU44 | Lift the lower eyelid up |
| AU11 | Deepen the middle nasolabial fold | AU23 | Tighten the lips into a line | AU45 | Blink two eyes |
| AU12 | Raise the corners of mouth | AU24 | Squeeze the lips together | AU46 | Blink one eye |

the effect of facial expression recognition based on LDA feature dimensionality reduction, including traditional artificial features, deep learning features, and the hybrid features. First, multidimensional feature extraction is performed. For traditional artificial features, we extract face AU information according to the facial action coding system and construct a 35-dimensional AU feature set. For deep learning features, we use the facial feature extractor based on FaceNet pretraining to extract the deep facial features and construct a 128-dimensional feature set. Second, LDA dimension reduction is performed on the extracted features respectively. Then, machine learning models, such as Naive Bayes and decision trees, are used to analyze the effects of facial expression recognition before and after LDA feature dimensionality reduction. Finally, we compare the effects of PCA and LDA feature dimensionality reduction methods on facial expression recognition. The specific workflow is shown in Figure 1.

3. Methods

3.1. Face Multidimensional Feature Extraction. In order to fully express the face features, we use different feature extraction methods to extract two aspects of face features according to the KDEF dataset: one is the AU feature based on the artificial methods and the other is the deep feature based on the deep learning.

3.1.1. Face AU Feature Extraction. We adopt the method of AU feature extraction in OpenFace [18] to extract the AU features in face images. As listed in Table 1, the subsets of AU that OpenFace can recognize are as follows: AU01, AU02, AU04, AU05, AU06, AU07, AU09, AU10, AU12, AU14, AU15, AU17, AU20, AU23, AU25, AU26, AU28, and AU45.

OpenFace describes the recognizable AUs in two scoring methods: (1) existence: if the AU is visible on the face, then 0 is absent and 1 is present; (2) intensity: if the AU has the intensity score on the interval $[0, 5]$, then 0 is absent and 5 is present at the maximum intensity. OpenFace is able to judge

the existence of 18 recognizable AU subsets and score the intensity of 17 AUs except for AU28. Therefore, each face image can form 35-dimensional AU features, as shown in Figure 2.

3.1.2. Face Deep Feature Extraction. In order to enable the deep learning model to focus on the face and better characterize the face, we use the method described in literature [10] to extract the key points from the dataset and correct the face according to the key points. It uses the gradient boosting decision tree (GBDT) regression method to accurately identify the key points of face. One GBDT is composed of multiple trees. Through these trees, the initial face key points are gradually returned to the real face key points. The specific implementation steps are as follows:

- (1) There are generally two ways to select the initial face key points. The first one is to take the mean value of real key points of all images as the initial face key points. The second one is to randomly select the real key points of other images as the initial key points of the current image. This study chooses the latter method to select the initial face key points.
- (2) When a tree is constructed and multiple images are input into the current tree, each image will fall into a leaf node. Calculate the differences between the current face key points and the real face key points of each image, and take the mean value of the differences of all images in the same leaf node as the residual saved by the leaf node.
- (3) Update the current face key points of each image as the sum of the current face key points and the residual.
- (4) Repeat steps 2 and 3 until enough trees are established. In that way, the last current face key points can be used as the real face key points.

According to the above algorithm, we extract 68 face key points for each face image and use these key points to

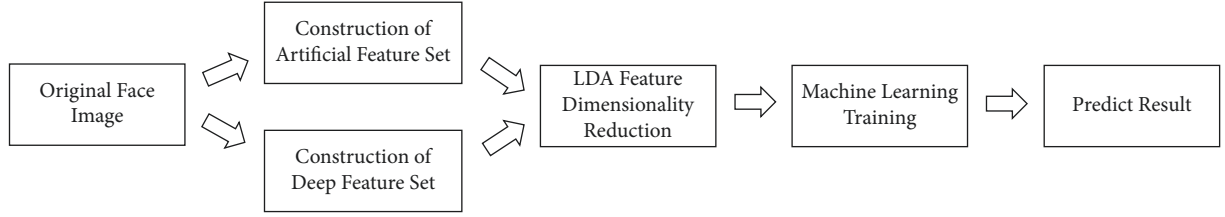


FIGURE 1: Workflow of facial expression recognition.



FIGURE 2: AU feature extraction of face.

perform the affine transformation correction on the face, so that the corners of the eyes and the nose are close to the horizontal position. It is shown in Figure 3.

After correcting the face image, we use the face feature extractor in OpenFace to extract deep face features. This feature extractor is based on the pretraining deep learning model FaceNet [19], which is shown in Figure 4.

OpenFaces uses 500,000 images to train FaceNet, including the two largest labeled face recognition datasets, CASIA-WebFaces [20], and FaceScrib [21]. It can extract the 128-dimensional deep feature of efficient facial feature expression.

3.2. Face Feature Dimensionality Reduction. Due to the diversity of facial expressions, recognizing facial expressions is a multiclassification task, so it needs to use the multiclassification promotion form of LDA. For the problem of multiclassification, we assume the dataset c , in which any sample x_i is an n -dimensional vector, and y_i is its category, $y_i \in \{n_1, n_2, \dots, n_k\}$, with a total of k categories. N_j ($j = 1, 2, 3, \dots, k$) is defined as the number of Class j sample, X_j ($j = 1, 2, 3, \dots, k$) is defined as the set of Class j sample, μ_j ($j = 1, 2, 3, \dots, k$) is defined as the mean vector of

Class j sample, and \sum_j ($j = 1, 2, 3, \dots, k$) is defined as the generalized covariance matrix of Class j sample.

Assuming that the matrix formed by the projected hyper-plane basis is $W = (w_1, w_2, \dots, w_d)$, then the optimization objective is as follows:

$$\arg \max_w J(W) = \frac{w^T S_b w}{w^T S_w w}, \quad (1)$$

where S_w is the intraclass scatter matrix:

$$S_w = \sum_{j=1}^k \sum_j = \sum_{j=1}^k S_{w_j} = \sum_{j=1}^k \sum_{x \in X_j} (x - \mu_j)(x - \mu_j)^T. \quad (2)$$

S_b is the interclass scatter matrix:

$$S_b = \sum_{j=1}^k N_j (\mu_j - \mu)(\mu_j - \mu)^T. \quad (3)$$

In equation (3),

$$\mu = \frac{1}{m} \sum_{i=1}^m x_i. \quad (4)$$

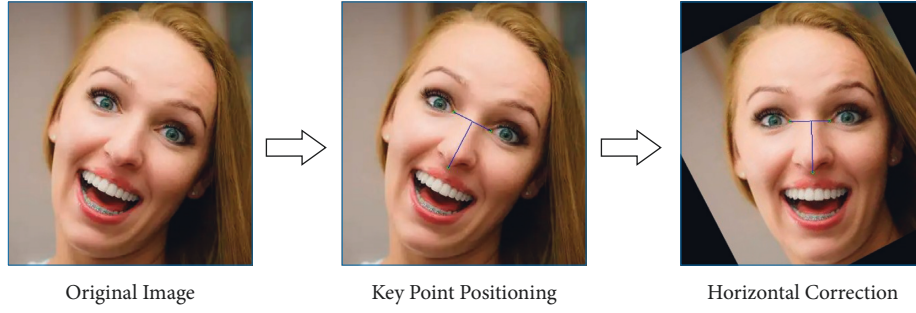


FIGURE 3: Process of face correction.

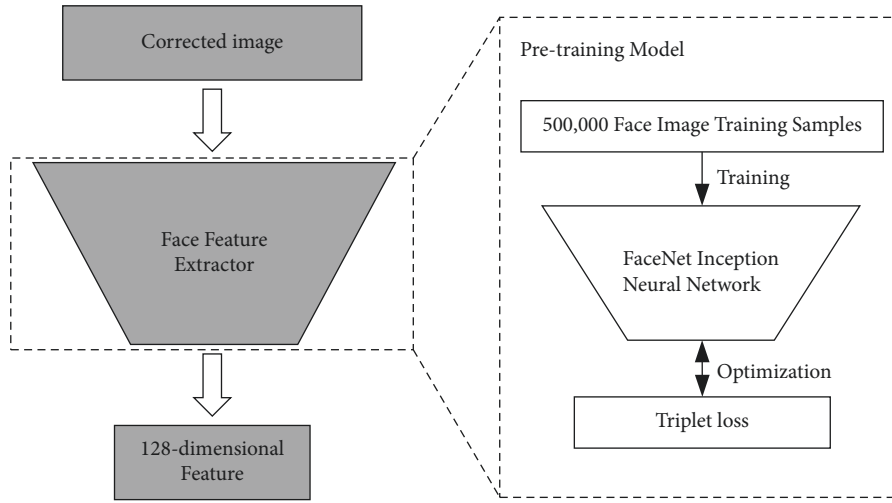


FIGURE 4: 128-dimensional deep feature extraction of face.

Since $J(W)$ is a matrix, the diagonal element product is used to replace the matrix for optimization:

$$\hat{J}(W) = \frac{\prod_{i=1}^d w_i^T S_b w_i}{\prod_{i=1}^d w_i^T S_w w_i} = \prod_{i=1}^d \frac{w_i^T S_b w_i}{w_i^T S_w w_i}. \quad (5)$$

In equation (5), the right side of $\hat{J}(W)$ is the continuous multiplication form of the generalized Rayleigh quotient. It only needs to obtain the maximum value one by one and then $J(W)$ will obtain the maximum value. As the maximum value of the Rayleigh quotient $w^T S_b w / w^T S_w w$ is the maximum eigenvalue of $S_w^{-1} S_b$, the matrix W formed by the eigenvectors corresponding to the product of the largest d eigenvalues is the mapping result after feature dimensionality reduction.

3.3. Introduction of Machine Learning Classification Models.

In order to evaluate the influence of the feature set on the face recognition after LDA dimensionality reduction, we rely on the several machine learning methods for facial expression recognition, including Naive Bayes classification model (NB), the decision tree classification model (DT), K-nearest neighbor (KNN), random forest (RF), and support vector machine (SVM). Among them, the Naive Bayes model [22] outputs the probability that a sample belongs to a specific category by predicting the membership probability

of the samples and each category. It is assumed that $S = (s_1, s_2, \dots, s_n)$ is an item to be classified and each s_i is a feature attribute of S . According to the existing category set $C = (c_1, c_2, \dots, c_m)$, we can calculate the probability of all categories under the characteristic condition of S and select the category with the highest probability as the category label of S . In this way, we can obtain the equation definition of Naive Bayes classifier as shown in the following equation:

$$V(S) = \arg \max_{c_i \in C} \frac{p(c_i) \prod_{k=1}^n p(s_k | c_i)}{p(S)}. \quad (6)$$

In equation (6), $V(S)$ is the category label of S , $p(S)$ is a constant for all categories, and $p(c_i)$ is the category prior probability. $p(s_1 | c_i), p(s_2 | c_i), \dots, p(s_n | c_i)$ is the conditional probability of each feature attribute in S under the condition of category c_i . All of these can be obtained from the training sample sets.

The decision tree classification model [23] consists of three parts: root node, internal node, and leaf node. The root node is the starting point of decision-making, and a decision path can be formed from the root node to a leaf node. Each of the internal nodes contains a feature that can divide the data. Each leaf node corresponds to a category. We use the classic algorithm ID3 to build a decision tree model. We suppose that S is the set of training samples, $|S|$ is the number of training samples, and A is any feature of the sample.

When the samples are divided into n different classes C_1, C_2, \dots, C_n , and the number of samples in these classes is marked as $|C_1|, |C_2|, \dots, |C_n|$, then the probability that any sample S belongs to class C_i is as follows:

$$p(S_i) = \frac{|C_i|}{|S|}. \quad (7)$$

Then, the total information entropy is as follows:

$$\text{Entropy}(S) = - \sum_{|S|} \sum_n p(S_i) \log p(S_i). \quad (8)$$

The information entropy of samples divided in accordance with Feature A is as follows:

$$\text{Entropy}(S, A) = \sum_A \left(\frac{|S_v|}{|S|} \right) \text{Entropy}(S_v). \quad (9)$$

The information gain of Feature A on the set S is as follows:

$$\text{Gain}(S, A) = \text{Entropy}(S) - \text{Entropy}(S, A). \quad (10)$$

In equation (9), \sum_A is the sum of all possible values v of attribute A , S_v is the subset of sample features A with v value in S , and $|S_v|$ is the number of samples in S_v . According to the theory of information theory, the ID3 algorithm uses ‘‘information gain’’ to measure uncertainty; that is, the greater the information gain, the smaller the uncertainty and the better the division effect.

4. Experimental Results and Analysis

The experimental conditions in this study are CPU: Intel(R) Core (TM) i5-10210U, RAM: 16.0 GB, and the programming language is Python 3.6.

4.1. Datasets and Evaluation Metrics. The experiments in this study are based on the KDEF public face dataset [24]. The KDEF dataset was originally used for psychological and medical experiments on perception, attention, emotion, memory, etc. It contains the facial information of 70 people, 35 males and 35 females, who are between the ages of 20 and 30. They have no beards, earrings, or glasses and have removed their makeup as far as possible before shooting. In the process of collecting the dataset, all the subjects wore special gray T-shirts, and the light was soft indirect light, evenly distributed on both sides of the face. Seven different expressions are collected for each person, including surprise, sadness, happiness, etc., and each expression has the color images from 5 angles. The entire dataset has a total of 4900 images, each with a size of $562 * 762$ pixels. Among them, there are 2916 photos that can identify the frontal and half-side face. Some samples are shown in Figure 5.

We randomly select 10% of the identifiable samples as the test set and the rest as the training set. The recognition effect is measured by the multiclassification evaluation metrics, including microprecision (micro-P), microrecall (micro-R), micro-F1, macroprecision (macro-P), microrecall

(macro-R), and macro-F1, which is shown in equations (11)–(16), where N denotes the number of categories, TP is true positive, FP is false positive, and FN is false negative.

$$\text{micro-Precision} = \frac{\sum_{i=1}^N TP_i}{\sum_{i=1}^N TP_i + \sum_{i=1}^N FP_i}, \quad (11)$$

$$\text{micro-Recall} = \frac{\sum_{i=1}^N TP_i}{\sum_{i=1}^N TP_i + \sum_{i=1}^N FN_i}, \quad (12)$$

$$\text{micro-F1} = 2 \frac{\text{micro-Recall} \times \text{micro-Precision}}{\text{micro-Recall} + \text{micro-Precision}}, \quad (13)$$

$$\text{macro-Precision} = \frac{1}{N} \sum_{i=1}^N \frac{TP_i}{TP_i + FP_i}, \quad (14)$$

$$\text{macro-Recall} = \frac{1}{N} \sum_{i=1}^N \frac{TP_i}{TP_i + FN_i}, \quad (15)$$

$$\text{macro-F1} = \frac{1}{N} \sum_{i=1}^N F1_i. \quad (16)$$

4.2. Analysis of LDA Dimensionality Reduction Effect for Facial Expression Recognition. In order to test the effect of LDA feature dimensionality reduction on the accuracy of facial expression recognition, we respectively extract the AU feature and the 128-dimensional feature from the images in the dataset as two different feature sets and regard the hybrid feature set as the third feature set. For the three feature sets, LDA was used to reduce the dimensions to 7 dimensions, and the training set of the machine learning method is used to test its classification effect. Therefore, the experiment can be divided into three parts: analysis of LDA dimensionality reduction effect based on AU features, analysis of LDA dimensionality reduction effect based on 128-dimensional face deep features, and analysis of LDA dimensionality reduction effect based on hybrid features.

4.2.1. Analysis of LDA Dimensionality Reduction Effect Based on AU Features. The facial AU feature is obtained by dividing the face into regions and counting the influence of each expression on different regions. It contains rich expression information, but it may also have redundancy that affects the recognition results. Therefore, we adopt several machine learning methods for facial expression recognition on basis of the AU feature and compare the recognition effects before and after LDA feature dimensionality reduction. The results are listed in Table 2.

Macro-F1 is taken as an example in Table 2. After the LDA dimensionality reduction, the macro-F1 of the NB and DT increased, while that of KNN, RF, and SVM decreased. This shows that there is some redundancy in the original AU feature set.

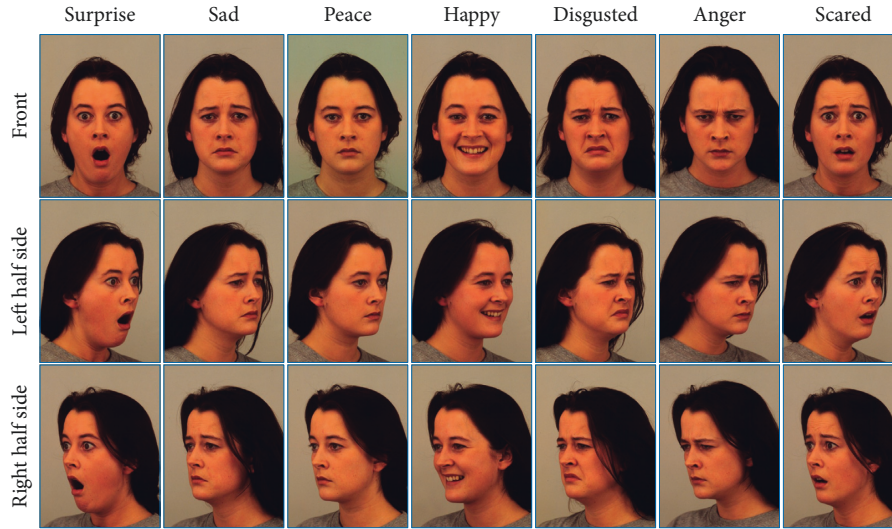


FIGURE 5: Samples in the KDEF.

TABLE 2: LDA dimensionality reduction effect of AU feature.

| Method | NB | | DT | | KNN | | RF ($n = 100$) | | SVM | |
|----------|--------|--------|--------|--------|--------|--------|------------------|--------|--------|--------|
| | Before | After | Before | After | Before | After | Before | After | Before | After |
| Micro-P | 0.5833 | 0.6756 | 0.5568 | 0.5667 | 0.6604 | 0.6517 | 0.6956 | 0.6699 | 0.7038 | 0.6815 |
| Macro-P | 0.5446 | 0.6731 | 0.5536 | 0.5640 | 0.6535 | 0.6527 | 0.6975 | 0.6724 | 0.7054 | 0.6840 |
| Micro-R | 0.5468 | 0.6743 | 0.5546 | 0.5650 | 0.6553 | 0.6541 | 0.6991 | 0.6737 | 0.7071 | 0.6854 |
| Macro-R | 0.5446 | 0.6731 | 0.5536 | 0.5640 | 0.6535 | 0.6527 | 0.6975 | 0.6724 | 0.7054 | 0.6840 |
| Micro-F1 | 0.5644 | 0.6749 | 0.5557 | 0.5659 | 0.6578 | 0.6529 | 0.6974 | 0.6718 | 0.7054 | 0.6834 |
| Macro-F1 | 0.5446 | 0.6731 | 0.5536 | 0.5640 | 0.6535 | 0.6527 | 0.6975 | 0.6724 | 0.7054 | 0.6840 |

4.2.2. Analysis of LDA Dimensionality Reduction Effect Based on 128-Dimensional Deep Features. The feature extraction based on the deep learning model can perform a higher-level abstract expression. In contrast to the local AU features, the semantic information of these features is richer, but the detailed location information is relatively brief. The facial expression recognition effect of 128-dimensional deep features and the dimensionality reduction effect of LDA are listed in Table 3.

It can be seen from Table 3 that the macro-F1 of facial expression recognition based on the original face 128-dimensional deep features is low. After LDA dimensionality reduction, the macro-f1 of all the methods is significantly improved except for the RF method.

The reason is that most of the original 128-dimensional deep features are facial semantic information, which has little effect on the facial expression recognition that requires detailed information. Therefore, the recognition effect before dimensionality reduction is low. LDA can project the feature to the largest distance between classes according to the label, which can also be regarded as mapping the 128-dimensional deep features to the feature space that can conduct the facial expression recognition. Due to the high expressive ability of deep features, these features can better express facial expression information after dimensionality reduction. Therefore, compared with face AU features, the 128-dimensional deep features can achieve higher recognition accuracy after dimensionality reduction.

4.2.3. Analysis of LDA Dimensionality Reduction Effect Based on Hybrid Features. We spliced local face AU features and 128-dimensional deep features containing more global semantic information to obtain the hybrid features. The facial expression recognition effect of hybrid features is listed in Table 4.

As listed in Table 4, compared with the above single feature set, the hybrid features can achieve higher recognition accuracy before and after LDA dimensionality reduction. After dimensionality reduction, the macro-F1 of the NB, DT, KNN, RF, and SVM methods increased by 19.22%, 18.04%, 15.10%, 3.91%, and 11.83%, respectively. The reason is that the hybrid features contain multidimensional face features, namely, local face AU features and global 128-dimensional deep semantic features. It enables the model to be fully learned, so as to achieve the higher recognition accuracy.

Combining Tables 2–4, it can be seen that no matter which feature set is used, LDA dimensionality reduction can effectively improve the recognition accuracy of the two machine learning methods. The main reason is that LDA feature dimensionality reduction increases the correlation between features and categories; that is, the class spacing is maximized while the intraclass dispersion is minimized, so that the subsequent model can better learn the interclass differences, resulting in the higher recognition evaluation matrix. The dimensionality reduction efficiencies of LDA for

TABLE 3: LDA dimensionality reduction effect of 128-dimensional deep features.

| Method | NB | | DT | | KNN | | RF ($n = 100$) | | SVM | |
|----------|--------|--------|--------|--------|--------|--------|------------------|--------|--------|--------|
| | Before | After | Before | After | Before | After | Before | After | Before | After |
| Micro-P | 0.4459 | 0.7075 | 0.3722 | 0.5903 | 0.6557 | 0.6965 | 0.7642 | 0.6986 | 0.6336 | 0.7121 |
| Macro-P | 0.4439 | 0.7067 | 0.3709 | 0.5889 | 0.6131 | 0.6939 | 0.7644 | 0.6983 | 0.6309 | 0.7119 |
| Micro-R | 0.4442 | 0.7076 | 0.3715 | 0.5894 | 0.6141 | 0.6948 | 0.7664 | 0.6993 | 0.6328 | 0.7127 |
| Macro-R | 0.4439 | 0.7067 | 0.3709 | 0.5889 | 0.6131 | 0.6939 | 0.7644 | 0.6983 | 0.6309 | 0.7119 |
| Micro-F1 | 0.4450 | 0.7075 | 0.3719 | 0.5899 | 0.6342 | 0.6957 | 0.7653 | 0.6989 | 0.6332 | 0.7124 |
| Macro-F1 | 0.4439 | 0.7067 | 0.3709 | 0.5889 | 0.6131 | 0.6939 | 0.7644 | 0.6983 | 0.6309 | 0.7119 |

TABLE 4: LDA dimensionality reduction effect of hybrid features.

| Method | NB | | DT | | KNN | | RF ($n = 100$) | | SVM | |
|----------|--------|--------|--------|--------|--------|--------|------------------|--------|--------|--------|
| | Before | After | Before | After | Before | After | Before | After | Before | After |
| Micro-P | 0.6725 | 0.8212 | 0.5570 | 0.7374 | 0.6649 | 0.8089 | 0.7740 | 0.8143 | 0.7089 | 0.8296 |
| Macro-P | 0.6283 | 0.8205 | 0.5545 | 0.7349 | 0.6576 | 0.8086 | 0.7748 | 0.8139 | 0.7107 | 0.8290 |
| Micro-R | 0.6298 | 0.8213 | 0.5558 | 0.7357 | 0.6595 | 0.8096 | 0.7764 | 0.8147 | 0.7123 | 0.8298 |
| Macro-R | 0.6283 | 0.8205 | 0.5545 | 0.7349 | 0.6576 | 0.8086 | 0.7748 | 0.8139 | 0.7107 | 0.8290 |
| Micro-F1 | 0.6505 | 0.8212 | 0.5564 | 0.7341 | 0.6622 | 0.8092 | 0.7752 | 0.8145 | 0.7106 | 0.8297 |
| Macro-F1 | 0.6283 | 0.8205 | 0.5545 | 0.7349 | 0.6576 | 0.8086 | 0.7748 | 0.8139 | 0.7107 | 0.8290 |

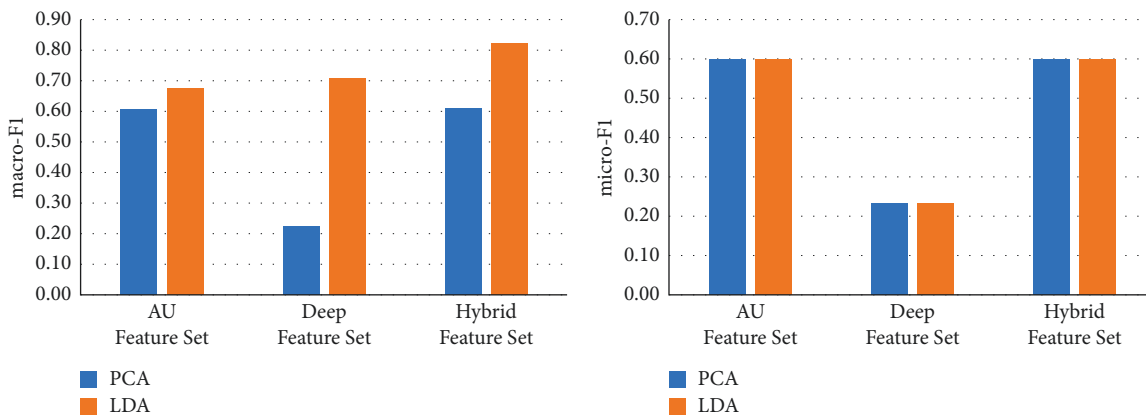


FIGURE 6: Comparison of the effects of different feature dimensionality reduction methods based on the NB.

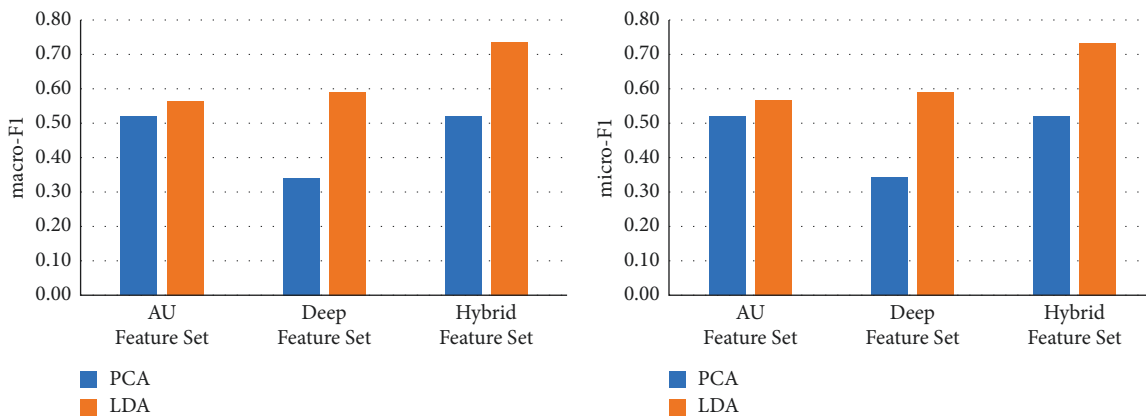


FIGURE 7: Comparison of the effects of different feature dimensionality reduction methods based on the DT.

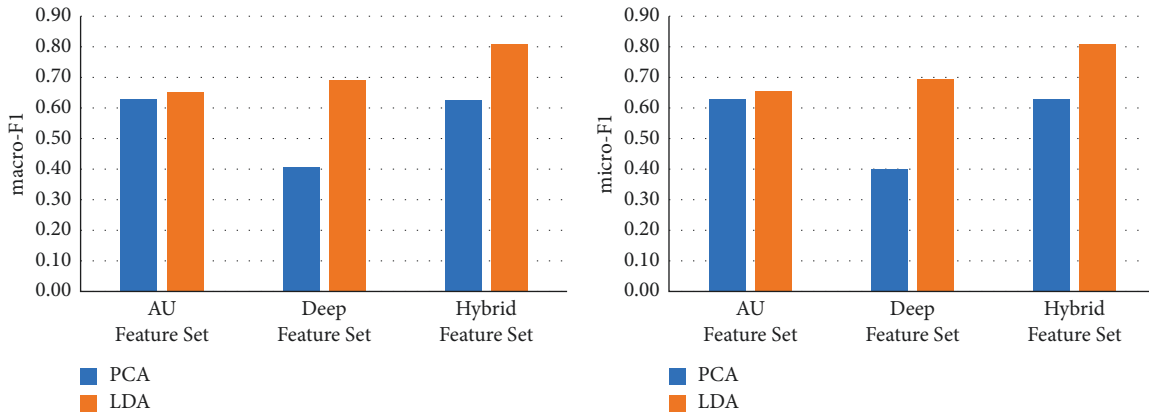


FIGURE 8: Comparison of the effects of different feature dimensionality reduction methods based on the KNN.

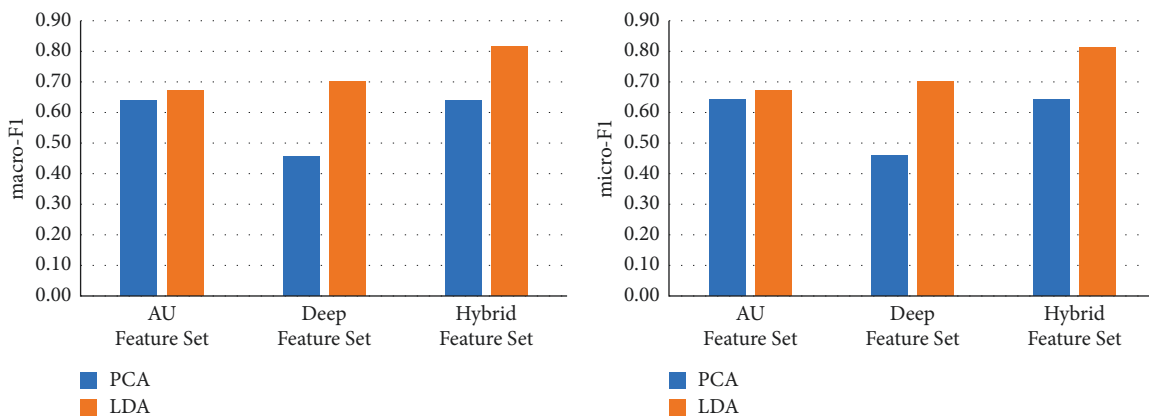


FIGURE 9: Comparison of the effects of different feature dimensionality reduction methods based on the RF.

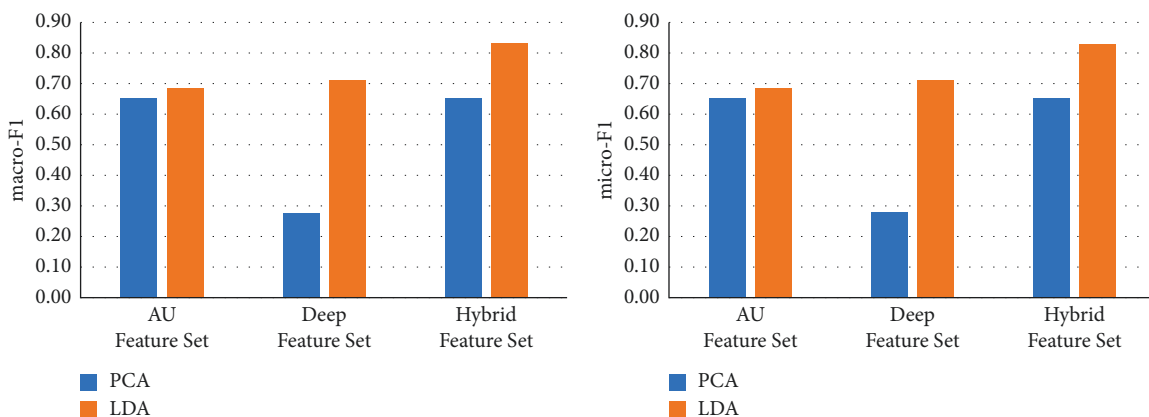


FIGURE 10: Comparison of the effects of different feature dimensionality reduction methods based on the SVM.

AU features, 128-dimensional deep features, and hybrid features are 15.14 s, 29.85 s, and 224.3 s, respectively.

4.3. Comparison of Different Dimensionality Reduction Methods. Principle component analysis (PCA) has been widely used as a classic method of unsupervised dimensionality reduction. Its main idea is to use the covariance matrix of the data to calculate the eigenvalues. The

eigenvalues are arranged from large to small, and the feature vectors corresponding to several largest representative components (features) are used to represent the overall characteristics of the data, so as to achieve the purpose of feature dimensionality reduction. We use LDA and PCA respectively to reduce the feature dimension for different feature sets and keep the same feature dimension after dimensionality reduction. We use the dimensionality-reduced features for the training of machine learning models to test

the effect on face recognition. The results are shown in Figures 6–10.

It can be seen from Figures 6–10 that for different types of feature sets and different machine learning methods, the accuracy of LDA feature dimensionality reduction for face recognition is higher than that of PCA method. Furthermore, the standard deviation of the correct samples identified by LDA is generally lower than that of PCA method, which indicates that LDA feature dimensionality reduction can make the feature be of greater category expression ability. Combining the ACC without dimensionality reduction in Tables 2–4, it can be seen that the accuracy of face recognition after PCA dimensionality reduction is even lower than that before dimensionality reduction. The reason is that PCA is an unsupervised data dimensionality reduction method. As this method does not use category information, it cannot obtain the effective dimensionality reduction effect for complex face feature sets. LDA is a supervised data dimensionality reduction method. It can be label oriented and select the vector that makes the Fisher criterion function reach the extreme value as the best projection direction, so that the sample can achieve the largest interclass dispersion and the smallest intraclass dispersion after being projected in this direction.

5. Conclusion and Discussion

The accurate recognition of facial expression recognition is inseparable from the extraction of multidimensional features, but there is often a lot of redundancy in multidimensional features. Based on the KDEF face public dataset, this study evaluates the effect of LDA feature dimensionality reduction method on facial expression recognition. First, the extraction of multitype features is performed on the face, including 35-dimensional face AU features based on artificial construction, 128-dimensional face deep features based on the FaceNet pretraining model, and 163-dimensional hybrid features. Second, LDA was used to reduce the dimensions of three different feature sets to 7 dimensions. Then, the classical machine learning methods are used to evaluate the effect of facial expression recognition before and after feature dimensionality reduction. Finally, we compare the LDA and PCA feature dimensionality reduction methods. The experimental results show that after the LDA feature dimensionality reduction, the facial expression recognition effects of the three feature sets have been improved to a certain extent, which are better than the PCA method. The dimensionality reduction of hybrid features has the best effect on improving the effect of facial expression recognition, with the macro-F1 reaching 82.90%. In addition, we found that no matter whether the feature dimensionality reduction is carried out or what machine learning method is used, the hybrid features can achieve a higher recognition accuracy than a single one. It shows that it is important to extract the multidimensional hybrid features in facial expression recognition.

As this study only compares the same number of dimensions when reducing the dimension, it may need to dynamically adjust the feature dimensions for different

feature sets. Hence, in the future, we will explore the effect of the feature dimensionality reduction under the optimal number of dimensions. The research on the effect of feature dimensionality reduction will be further expanded from multiple directions, such as more machine learning methods, more feature dimensionality reduction methods, and more different types of datasets.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

- [1] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition," *IEEE Transactions on Image Processing*, vol. 11, no. 4, pp. 467–476, 2002.
- [2] T. Ahonen, A. Pietikainen, and M. Pietikainen, "Face description with local binary patterns: application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006.
- [3] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, "Local Gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition," *Tenth IEEE International Conference on Computer Vision (ICCV'05)*, vol. 1, pp. 786–791, 2005.
- [4] D. Chen, X. Cao, F. Wen, and J. Sun, "Blessing of dimensionality: high-dimensional feature and its efficient compression for face verification," *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 12, pp. 3025–3032, 2013.
- [5] E. Friesen and E. Paul, "Facial Action Coding System," *A Technique for the Measurement of Facial Movement* Palo Alto 3, vol. 12, 1978.
- [6] J. Hamm, C. G. Kohler, R. C. Gur, and R. Verma, "Automated facial action coding system for dynamic analysis of facial expressions in neuropsychiatric disorders," *Journal of Neuroscience Methods*, vol. 200, no. 2, pp. 237–256, 2011.
- [7] P. N. Belhumeur, D. W. Jacobs, D. J. Kumar, and N. Kumar, "Localizing parts of faces using a consensus of exemplars," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 12, pp. 2930–2940, 2013.
- [8] X. Cao, Y. Wei, F. Wen, and J. Sun, "Face alignment by explicit shape regression," *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 2887–2894, 2012.
- [9] S. Karanwal and M. Diwakar, "OD-LBP: orthogonal difference-local binary pattern for Face Recognition," *Digital Signal Processing*, vol. 110, no. 10, Article ID 102948, 2021.
- [10] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 3, pp. 1867–1874, 2014.
- [11] R. Bro and A. K. Smilde, "Principal component analysis," *Analytical Methods*, vol. 6, no. 9, pp. 2812–2831, 2014.
- [12] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, vol. 7, no. 2, pp. 179–188, 1936.
- [13] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces,"

- IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 103–108, 1990.
- [14] S. Mondal and S. Bag, “Face recognition using PCA and minimum distance classifier,” in *Proceedings of the 5th International Conference on Frontiers in Intelligent Computing: Theory and Applications*, pp. 397–405, NY City, July 2016.
 - [15] L. Zhuang and Y. Guan, “Improvement of face recognition with gabor, PCA, and SVM under complex illumination conditions,” *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 23, no. 3, pp. 465–473, 2019.
 - [16] M. Bodini, A. D’Amelio, G. Grossi, R. Lanzarotti, and J. Lin, “Single sample face recognition by sparse recovery of deep-learned LDA features,” *Advanced Concepts for Intelligent Vision Systems*, vol. 23, pp. 297–308, 2018.
 - [17] A. Benouareth, “An efficient face recognition approach combining likelihood-based sufficient dimension reduction and LDA,” *Multimedia Tools and Applications*, vol. 80, no. 1, pp. 1457–1486, 2021.
 - [18] B. Amos, B. Ludwiczuk, and M. S. Openface, *A General-Purpose Face Recognition Library with mobile Applications*. CMU-CS-16-118, CMU School of Computer Science, Tech. Rep, China, 2016.
 - [19] F. Schroff, D. Kalenichenko, and P. James, “Facenet. A unified embedding for face recognition and clustering,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 815–823, Boston, MA, USA, 07-12 June 2015.
 - [20] Yi Dong, Z. Lei, S. Liao, and Z. Stan, “Learning face representation from scratch,” 2014, <https://arxiv.org/abs/1411.7923>.
 - [21] H.-W. Ng and S. Winkler, “A data-driven approach to cleaning large face datasets,” *IEEE International Conference on Image Processing (ICIP)*, vol. 265, no. 265, p. 530, 2014.
 - [22] K.-Yu Chou and Y.-P. Chen, “Real-time and low-memory multi-faces detection system design with naive Bayes classifier implemented on FPGA,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 11, pp. 4380–4389, 2020.
 - [23] O. Mary Kenneth, S. A. Bashir, O. Aderiike Abisoye, and A. Danlami Mohammed, “Face morphing attack detection in the presence of post-processed image sources using neighborhood component analysis and decision tree classifier,” *Proceedings of the Information and Communication Technology and Applications*, vol. 24, pp. 340–354, 2020.
 - [24] D. Lundqvist, A. Flykt, and A. Hman, *The Karolinska Directed Emotional Faces – KDEF, CD ROM from Department of Clinical Neuroscience, Psychology section*, China, 1998.