Research Article

Optimum Feature Selection with Particle Swarm Optimization to Face Recognition System Using Gabor Wavelet Transform and Deep Learning

Sulayman Ahmed,[1] Mondher Frikha,[1] Taha Darwassh Hanawy Hussein,[2] and Javad Rahebi [3]

[1]*ENETCOM, Universite de Sfax, Tunisia*
[2]*Kirkuk University, Kirkuk, Iraq*
[3]*Department of Software Engineering, Istanbul Ayvansaray University, Istanbul, Turkey*

Correspondence should be addressed to Javad Rahebi; cevatrahebi@ayvansaray.edu.tr

In this study, Gabor wavelet transform on the strength of deep learning which is a new approach for the symmetry face database is presented. A proposed face recognition system was developed to be used for different purposes. We used Gabor wavelet transform for feature extraction of symmetry face training data, and then, we used the deep learning method for recognition. We implemented and evaluated the proposed method on ORL and YALE databases with MATLAB 2020a. Moreover, the same experiments were conducted applying particle swarm optimization (PSO) for the feature selection approach. The implementation of Gabor wavelet feature extraction with a high number of training image samples has proved to be more effective than other methods in our study. The recognition rate when implementing the PSO methods on the ORL database is 85.42% while it is 92% with the three methods on the YALE database. However, the use of the PSO algorithm has increased the accuracy rate to 96.22% for the ORL database and 94.66% for the YALE database.

1. Introduction

Face recognition has attracted a lot of interest in recent years [1]. It has become one of the main areas of study in machine vision, pattern recognition, and machine learning. In face recognition, the system selects a face that is more like the desired face according to the trained faces and considers it as the final answer.

Facial recognition was proposed in the 1960s. The first semiautomatic facial recognition system was produced by Woody Bledsoe, Helen Chan Kurt, and Charles Bisson [2]. However, the human face includes a number of details that have been used in many systems, such as artificial age classification [3, 4], facial identification [5], forecasting images and restoration apps [6, 7], description of gender and gestures [8], human-computer interaction (HCI), electronic consumer experience management and audience recording, and tracking of security cameras. Applications for face recognition include monitoring, forensic and medical apps, security applications, in the banks, detection of the person in international centers of transition, access control, and several different fields. Recently, facial recognition technologies were widely used in particular in areas needing strict security measures (airports, police stations, banks, sports fields, and surveillance of entry and exit from business companies).

Computer security is considered to be important in the world today [9]. Face recognition remains an important subject in computer vision sciences. This is because the current systems perform well in relatively controlled conditions but appear to fail when there are issues with facial images, for example, presenting a particular face that differs by various factors, such as variations in posture, position, occlusion,

lighting, make-up, and noise- and blur-induced image damage. Although researchers have developed many technologies, multiple different solutions have been attempted to address the problem of changing conditions of the environment. These conditions are the main challenges to facial recognition. Difficulties of the face recognition problem derive from the fact that the faces tend to be approximately similar in their most typical shape (i.e., the front view), and the variations between them are very slight. As a consequence, frontal images formalize a large concentration of the size of the image. This size makes it nearly difficult for typical pattern recognition methods to recognize correctly with a high degree of level of success [10]. Another concentration is the database images [11]. It must have sufficient information for effective face recognition, so the recognition must be possible when dealing with the test image. It is also difficult to determine if there is enough information in the stored images so that the relevant information can be extracted from the databases. Often, unnecessary information is also present in the images of the database, resulting in higher storage consumption and higher processing times. In addition, the optimal size of the images requires to be stored in the databases for effective results [12, 13]. The image size can be compressed to the required size and be stored in the databases. When the image size is compressed, there would be a loss of features, but large numbers of these images can be stored and transmitted through the network fast [14].

In this paper, we used Gabor wavelet transform for feature extraction and then for reducing the features. To find the best feature, the PSO method is used. For the recognition of a face, the deep learning method with 6 layers is used.

2. Literature Review

Facial recognition is currently divided into two general categories: appearance-based methods, which statistically process the face, and model-based methods that operate geometrically [15]. For face recognition [16], discriminative dictionary learning and sparse representation are used. In their method, the Gabor amplitude images are implemented by the bank of Gabor filter. Furthemore, the local binary pattern (LBP) is used for feature extraction [17]. Face recognition can be considered one of the most significant applications in the image processing domain [18]. However, illumination and pose invariant recognitions are still the most obvious problems. Viewpoint and illumination are vital to the efficiency of the recognition system because these two factors differ when face images are taken in an uncontrolled environment. Elastic bunch graph matching [19], one of the feature-based methods, has been known for a long time to be accomplished toward several factors such as illumination and viewpoint [18]. Their excessive susceptibility to feature extraction and measurement of the extracted features [20] are what make them unreliable. As a result, the dominant method in the literature is appearance-based methods.

Ahonen et al. proposed a face recognition model with native binary patterns (LBP) [5]. In their study, the point ensuring the robustness of their work is that the algorithm is not sensitive to light.

The fisherface [21] technique is one of the milestones for face recognition under variations. In linear discriminant analysis (LDA), interperson alteration is used optimally with large and intraperson alteration efficaciously small to construct a subspace [21]. Like the PCA [22], the main disadvantage of this technique is that the data space is a consideration of Euclidean. The method does not succeed as multimodal distributed face images when data points are located in a nonlinear subspace.

The sparse representation algorithm based on the Gabor feature is proposed by Yang and Zhang [23]. In their method, the SRC and Gabor features are combined. Using this technique, they improved the human face recognition rate and reduced the complexity of computation.

The deep learning approaches are investigated [24]. The cross-resolution face recognition scenario based on the deep learning method is performed [25]. They robustly extracted the features by deep properties with a cross-resolution scenario. In [26], the angularly discriminative features based on deep learning for face recognition are utilized.

Xu et al. presented the new artificial neural network to face recognition called coupled autoencoder networks (CAN). This helps to overcome age-invariant face recognitions and redemption troubles [27].

The effect of variations in condition on face recognition has been investigated by authors [28]. Consequently, the dominant method has been the appearance-based method. Nikan and Ahmadi [29] introduced a new procedure that propped up fusion of global and local structures.

In [30], local linear transformations were used on behalf of one global transformation, which is a good improvement. The technique suggests different pose classes to different mapping functions. When a probe image is examined, its pose is determined by soft clustering. Deciding the number of pose clusters is a difficult task as in all clustering algorithms. Moreover, novel poses cannot be treated in case of critical variations. In [31], the authors used the neighborhood structure of the input space to determine the underlying nonlinear manifold of multimodal face images. What is used to calculate the basic set is called Laplacian Faces Locality Preserving Projections (LPP). When examining face images with other poses, facial expressions, and illumination conditions, their recognition performance was higher than that of fisherfaces or eigenfaces. In [32], pose variation using view-based eigenfaces was studied. For every view, eigenfaces were numbered to apply a standard dimensional subspace as separate transformations. In addition, a feature-based scheme is included within the eigenfeatures introduced by the authors. As in [33], their performance depends highly on decoupling. Here, the eigenilluminant field technique was used to identify the subspace of poses. Zhao et al. [34] prepared the blurry invariant binary identifier to face recognition. They enhanced the corral among the binary codes of sharp face images and blurred face images of positive image pairs about to learn matrix of projection. Then, they used the learned projection matrix to procure blur-robust binary codes by quantizing projected pixel difference vectors (PDVs) in the trial phase. The discriminative DL method by training a classifier of the coding coefficients is proposed
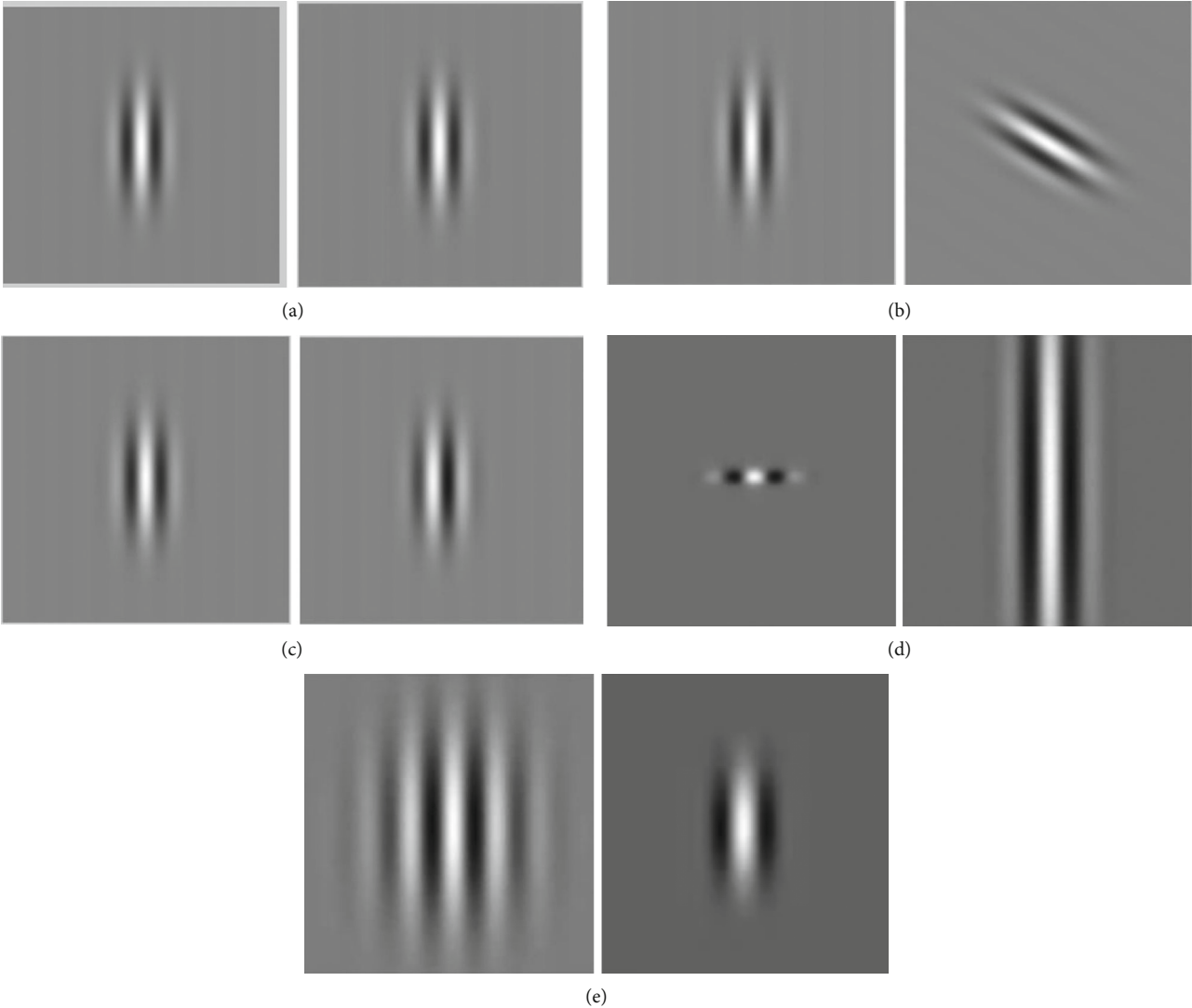
FIGURE 1: (a) Different values of wavelength (left: 25, right: 50), (b) different orientation (left: 0, right: 45), (c) changing the values of phase shift (left: 180, right: 90), (d) aspect ratio very large (left) and very small (right), and (e) different bandwidth values (left: large, right: small) [57].
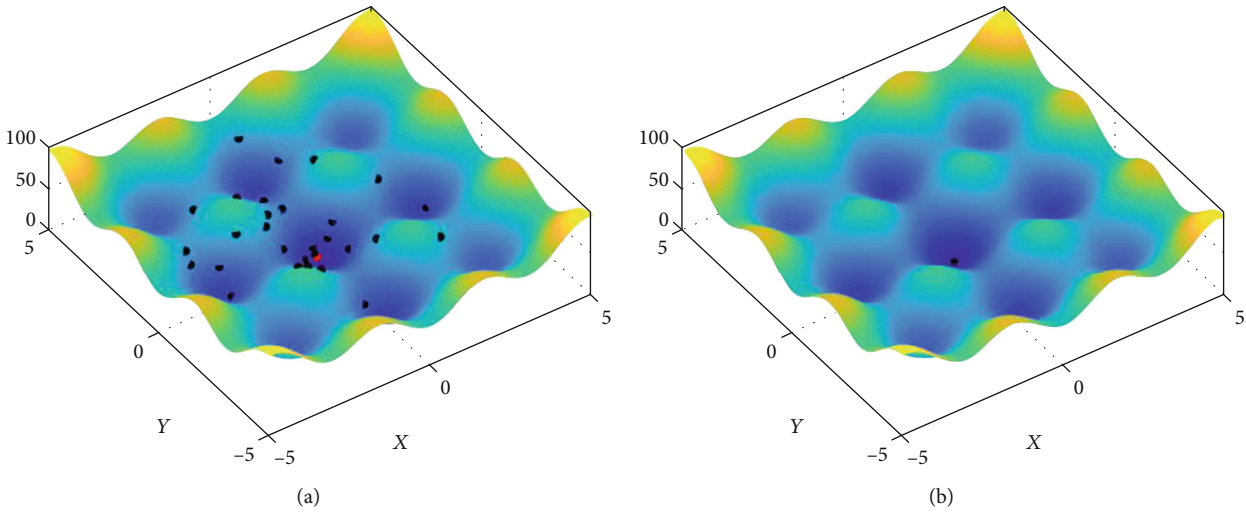

FIGURE 2: Particles searching for the best and optimum solution: (a) first iteration and (b) last iteration.

by Mairal et al. [35]. For texture classification and digit recognition, they verified their method. In [36], the sex and country population density interact to predict face recognition talent. The face plus word recognition based on the euro magnetic correlates of hemispheric specialization is presented in [37].

A method that is insensitive to illumination changes was produced by the authors in [38] through combining the generalized concept of photometric stereo and eigenlight field. 3D morphable face models were used in [20, 39, 40, 41] to defined novel poses, which have performances higher than that of the previous research works. Rendering ability for new poses and illumination conditions is exceptional with 3D morphable models [41]. However, the computational cost of generating 3D models from 2D images or using laser scanners to access 3D models decreases the efficiency of the recognition system.

Royer et al. [42] used the eye region to identify a face accurately. The mixed neighborhood topology with the cross decoded patterns is done by [43].

Illumination variance was studied in [44]. The quotient image was suggested by the authors as an identity signature that is insensitive to illumination. While the approximation does not work well, then the probe image has an unexpected shade. Its probe images could be identified with particular illumination then the gallery images. The technique requires only one gallery image for a thing. The technique in [45] introduced additional constraints on the albedo and the surface normal to solve the shadow problem. An illumination cone model was proposed in [20]. The authors discussed a series of images of an object in a fixed pose only describing a convex cone in all lighting conditions. The method needs some images to test their identity and then to guess its surface geometry and albedo map. They defined different illumination cones for each sampled viewpoint to deal with pose variations. The authors discussed in [46, 47] the use of all Lambert reflecting functions to create all kinds of illumination conditions for Lambertian objects. The researchers presented the approximation of plenty of variation of illumination achieved using only nine spherical harmonics. The multiple virtual views and alignment errors are presented in [48]. They manipulated the cross-pose face recognition method.

A methodology for recognition was also used in [46]. In [40], a spherical harmonics approach was exploited, and good recognition results were presented. They designed a 3D morphable model to achieve pose invariance, and this needs to generate 3D face models from 2D images.

Original and symmetrical examples of face training were used [49] to perform collaborative representation for face recognition.

A nonlinear subspace approach was introduced using the tensor representation of faces, such as facial expressions, illumination, and poses [50]. The n mode tensor Singular Value Decomposition (SVD) could form the basis of an image. In this technique, various images are required under different variations for each training identity. In [51], there was another nonlinear assumption for each identity in the database, and a gallery manifold is stored. When a test identity

TABLE 1: Parameter for particle swarm optimization to select the best features.

Parameter	Description	Value
N	Number of particles (population size)	40
T	Maximum number of iterations	15
c_1	Cognitive factor	3
c_2	Social factor	2.5
w_{max}	Maximum bound on inertia weight	0.8
w_{min}	Minimum bound on inertia weight	0.5
V_{max}	Maximum velocity	5

with several new poses needs to be defined, first, its probe manifold is constructed, then using manifold to manifold distance can help to define its identity.

The main drawback is the requirement of multiple images of the test person. The authors in [52] introduced a considerable idea by bilinear generative models to decompose orthogonal factors. They showed a separable bilinear mapping between the input space and the lower dimensional subspace. After determining all the parameters of mappings, identity and pose information can be separated explicitly. The recognition and synthesizing capabilities of the technique were analyzed, and the results were encouraging. In [53], illumination invariance was examined using a similar framework. In addition, a ridge regression technique was designed to come through the matrix inversion needed in the symmetric bilinear model. A modified asymmetric model in [54] is aimed at overcoming pose variations. One of the most important factors affecting performance is the solvation of the pose space. The authors in [55] incorporated the nonlinearity of the generative models. They recommended a nonlinear scheme combined with the bilinear model and tried to remove the linearity constraint of the classical generative models. Wright et al. [56] presented a robust method for face recognition. They used sparse representation for feature extraction.

3. Proposed Method

In this paper, the face recognition system undergoes stages. These three stages are feature extraction using Gabor wavelet transform, selecting the best features with the PSO method, and face recognition with the deep learning method.

3.1. Feature Extraction Using Gabor Wavelet Transform. A much useful instrument in image processing, especially in image identification, is the Gabor filter. The Gabor filter over the spatial field, which has two dimensions, is a Gaussian kernel function as explained below by a complex sinusoidal plane wave:

$$G(x, y) = \frac{f^2}{\pi \gamma \eta} \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \exp\left(j2\pi f x' + \phi\right).$$
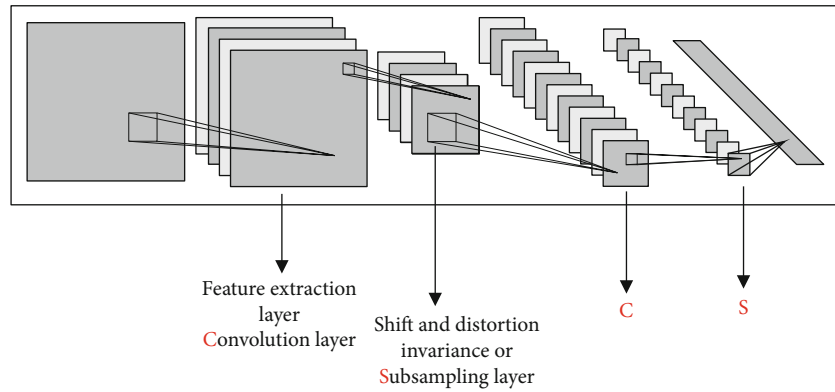
$$(1)$$

FIGURE 3: CNN's topology.

Here, f represents the sinusoid's frequency, θ is the orientation of the normal to the parallel Gabor function's stripes, ϕ is the phase offset, σ is the Gaussian envelope's standard deviation, and γ is the spatial aspect ratio that determines the elliptic support for the function of Gabor.

x' and y' can be calculated as the following equations:

$$x' = x \cos \theta + y \sin (\theta),$$
$$y' = -x \sin \theta + y \cos (\theta). \tag{2}$$

Figure 1 shows the influence of changing some parameters for Gabor's function.

Some of the various benefits of Gabor filters are invariance rotation, scaling, translation, and resistance to distortion of images such as illumination change [58, 59]. They are specially proper for fabric representation plus discrimination.

A range of Gabor filters with other frequencies and directions can be used to extract many features such as texture analysis and segmentation from an image [60]. By varying the orientation, we can look for fabric orientation in a specific direction. By varying the standard deviation of the Gaussian envelope, we change the basis' support or the image's size region being analyzed.

When the features are extracted, the best relative set of features are selected using the PSO method for a flexible face recognition system.

3.2. Feature Selection with Particle Swarm Optimization.

Particle swarm optimization (PSO) or known as the bird swarm algorithm was initially created in 1995 by Kenny and Eberhart [61]. PSO is a mathematical method that tries to solve optimization problems. For each problem, there are particles (solutions) flying over the problem area based on some mathematical calculations for the velocity and position of the particle. Each particle has fitness values that are measured by the fitness function to be optimized and has velocity that guides the flying of the particles [62].

In computational techniques, PSO is used as a random optimization algorithm for feature selection and classification. This is done by iteratively selecting the most relative and useful set of features to improve or maintain the classification performance for a robust facial recognition system [63].

FIGURE 4: ORL database.

The basic idea behind this algorithm is the coevolvement of different classes of birds rather than focusing on a certain class of birds. This algorithm contributes to effective search abilities [64]. The PSO algorithm is illustrated in Figure 2.

First, all the particles are assigned primary values; after that, fit values for each particles are estimated. Then, the current fit value is determined; if it is better than the previous one, then we upgrade it to the current value, but if the old fit value is better; we keep it [65]. The algorithm ends, and this process is repeated until the best solution is obtained.

The equation of the PSO algorithm is demonstrated below:

$$v_i^d(t+1) = wv_i^d(t) + c_1 r_1 \left(p\text{best}_i^d(t) - x_i^d(t) \right)$$
$$+ c_2 r_2 \left(g\text{best}^d(t) - x_i^d(t) \right). \tag{3}$$

Each particle is upgraded with two "best" values in each iteration. Here, v denotes velocity which is bounded between $w_{\max}$ and $w_{\min}$, w is inertia weight, and x is solution [66, 67].
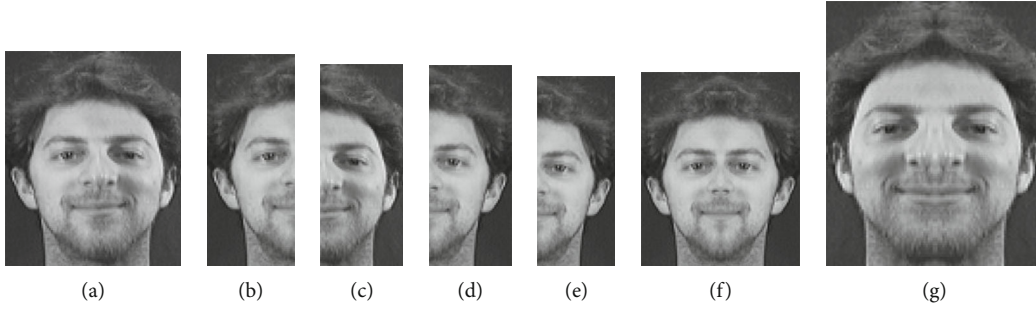
Figure 5: (a) Real image, (b) left side, (c) right side, (d) left side's mirror, (e) right side's mirror, (f) integration of left side with mirrors, and (g) integration of right side with mirrors.

Continuing, t refers to the number of irritation, i to the order of practicality in population, and d to the dimension of search space. c_1 and c_2 indicate acceleration factor; r_1 and r_2 are two independent random numbers in $[0, 1]$. $p$best implies the personal best solution (the best solution that has been found yet), while $g$best implies the global solution which is recorded by the particle swarm optimizer. This optimizer is the best worth yet achieved by any particle for the entire population.

Afterwards, velocity is updated to a probability value as demonstrated in the following equation:

$$s\left(v_i^d(t+1)\right) = \frac{1}{1 + e^{\left(-v_i^d(t+1)\right)}}. \tag{4}$$

Practical position and $p$best with $g$best are converted to the following equations:

$$x_i^d(t+1) = \begin{cases} 1, & \text{if } \text{rand} < s\left(v_i^d(t+1)\right), \\ 0, & \text{otherwise}, \end{cases} \tag{5}$$

where rand is a random number between 0 and 1.

$$p\text{best}_i(t+1) = \begin{cases} x_i(t+1), & \text{if } F(x_i(t+1)) < F(p\text{best}_i(t)), \\ p\text{best}_i(t), & \text{otherwise}, \end{cases}$$

$$g\text{best}_i(t+1) = \begin{cases} p\text{best}_i(t+1), & \text{if } F(p\text{best}_i(t+1)) < F(g\text{best}_i(t)), \\ g\text{best}_i(t), & \text{otherwise}, \end{cases}$$

$$\tag{6}$$

where F is the fitness function:

$$w = w_{\max} - (w_{\max} - w_{\min})\left(\frac{t}{T_{\max}}\right). \tag{7}$$

The parameter used for particle swarm optimization is shown in Table 1.

We obtained these parameters experimentally.

3.3. Convolutional Neural Network. The main component of a convolution neural network (CNN) is the convolution layer. The approach behind a convolution layer is a feature

Table 2: Specifications of the deep learning.

	Type	Activation	Learnable
1	Sequence input	5141	—
2	LSTM	200	Input weights ($800 * 5141$) Recurrent weights ($800 * 200$) Bias ($800 * 1$)
3	Fully connected	50	Weights $50 * 200$ Bias $50 * 1$
4	Dropout	50	—
5	Fully connected	1	Weights $1 * 50$ Bias $1 * 1$
6	Regression output	—	—

which has been learned locally for any given input (for example, any 2D images). It should be helpful in other regions of that same input source. For example, a feature for edge detection, which was proved useful in one part of the image, might be helpful in the other regions of the image at a possible general feature extraction stage. The learning of other features in an image such as edges oriented at an angle or curves is obtained by sliding the filters across the image with a step or stride size which is constant for a given convolution layer.

Layers of more than one subsampling and convolutional layer, preferably fully added layers, are called CNN. M is the height and width of the image, and r is the number of channels, while the input of an accessible layer is the image $m \times m \times r$, e.g., an RGB image has $r = 3$. The convolutional layer can differ in every core it has; this is because it will have k kernels or filters of size $n \times n \times q$, where n is much smaller than the size of the image and q could be smaller than the number of channels. Figure 3 shows the general topology of a CNN.

4. Experimental Result and Discussion

This chapter shows that the outcomes are derived from the simulation using MATLAB 2020a. The recognition system consists of three stages. The first is *the feature extraction*; in this stage, we used Gabor wavelet transform. The second is *feature selection*. In this stage, we used particle swarm
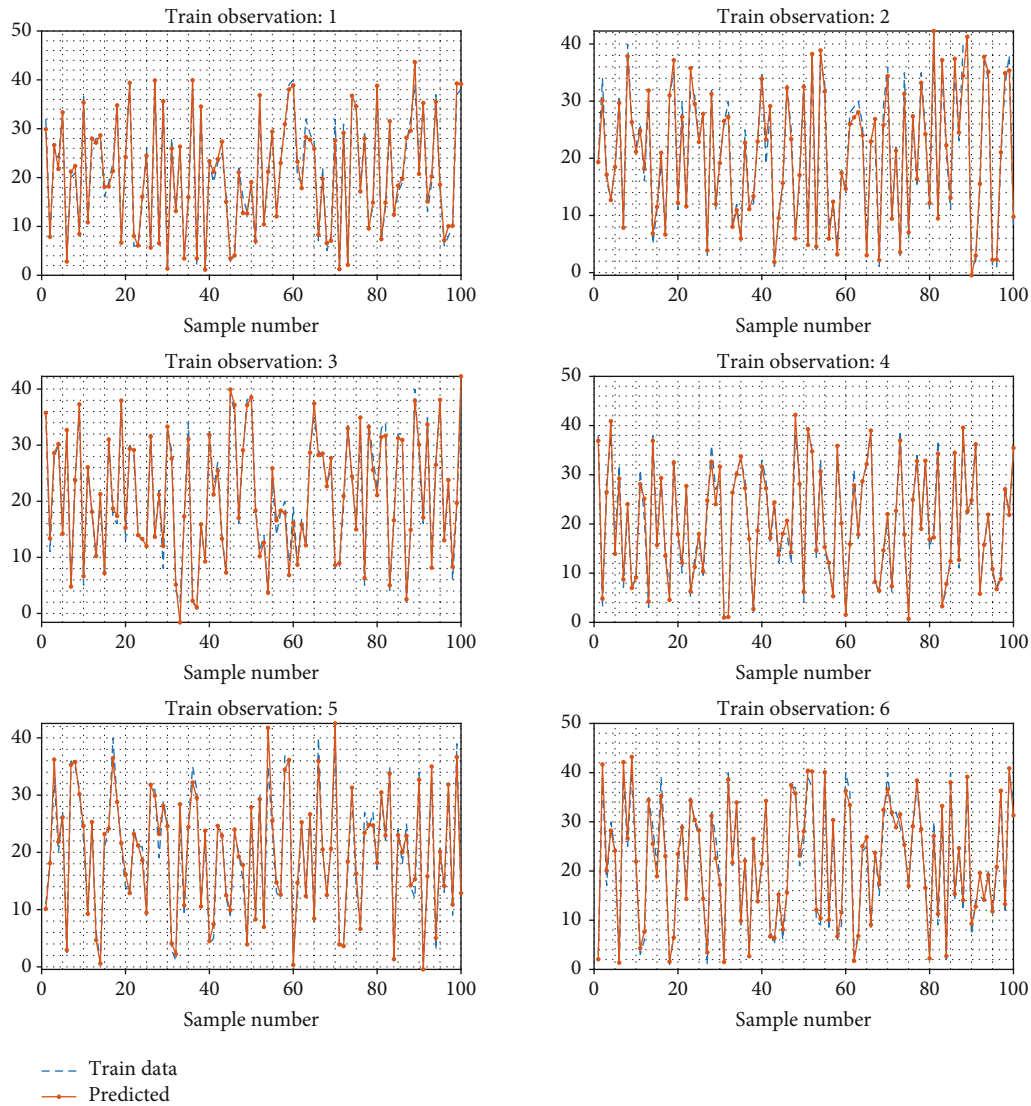
FIGURE 6: Train result.

optimization (PSO), on features that are obtained from Gabor wavelet transform. In the final stage, *the classification*, we used deep learning with 6 layers.

The database in this study is used from ORL databases. The ORL (Olivetti Research Laboratory) face database contains 400 images of 40 different people. There are ten different grayscale images of each of 40 distinct persons. Images were captured at various times, and they have various variations including various expressions (closed/open eyes, not smiling/smiling). The details of the face (with/without glasses) are included. Images were taken with a tolerance for some tilting and rotation of the face up to 20 degrees [49].

Some face images from the ORL database are shown in Figure 4.

Some simulation of the first face image is implemented on MATLAB 2020a, and the results are shown in Figure 5.

For evaluating the proposed method, we used the mean squared error (MSE), mean absolute percentage error
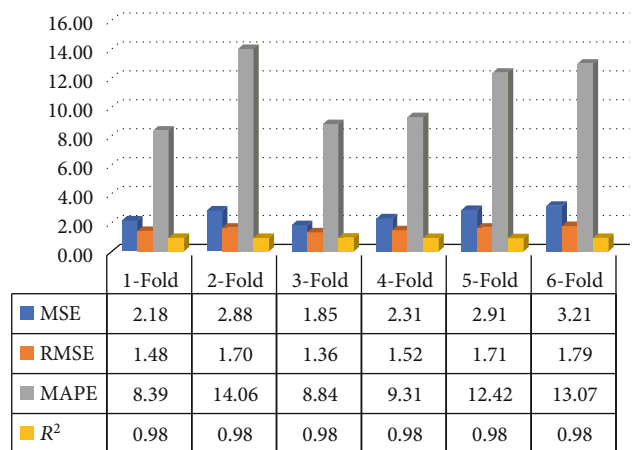

	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	6-Fold
MSE	2.18	2.88	1.85	2.31	2.91	3.21
RMSE	1.48	1.70	1.36	1.52	1.71	1.79
MAPE	8.39	14.06	8.84	9.31	12.42	13.07
R^2	0.98	0.98	0.98	0.98	0.98	0.98

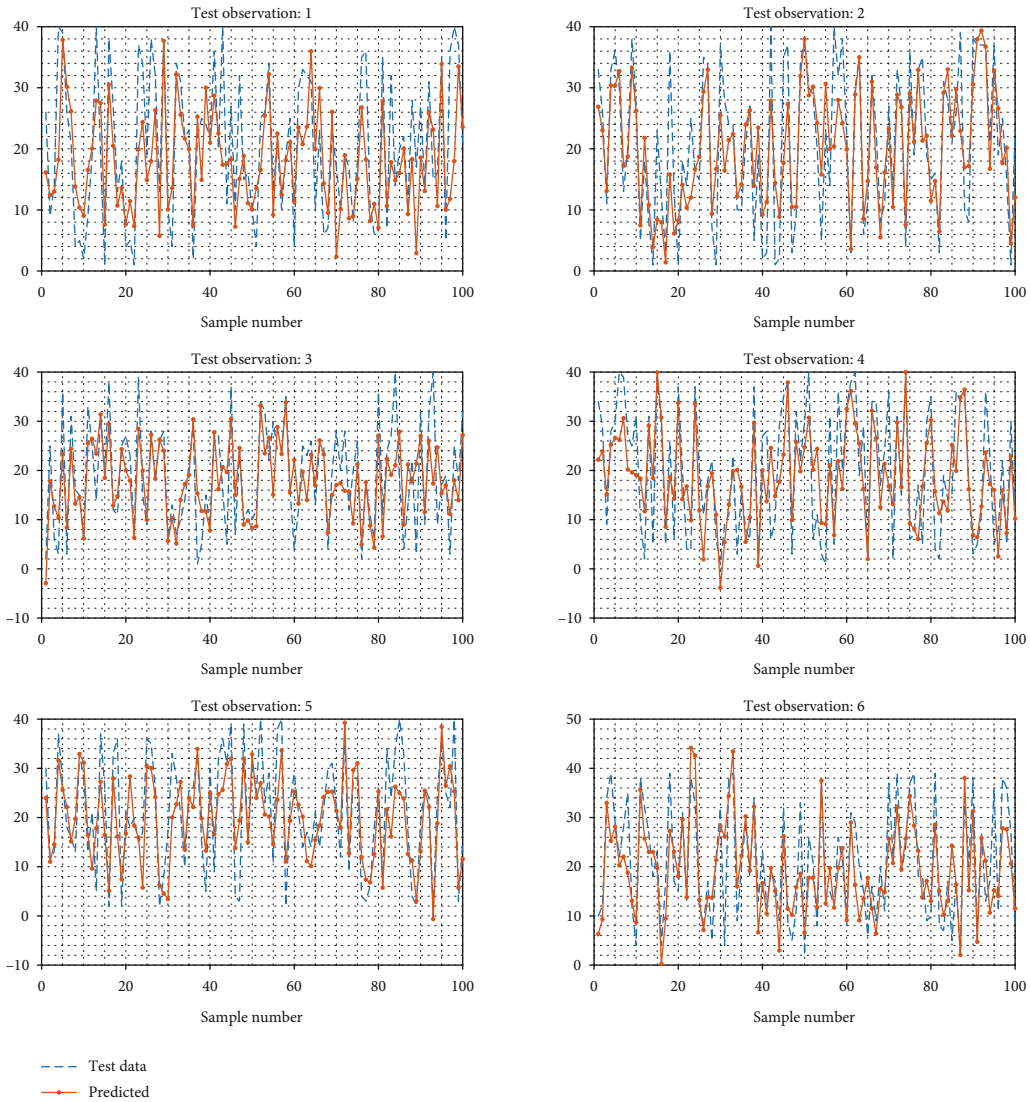FIGURE 7: The comparison of the MSE, RMSE, MAPE, and R for train data.

FIGURE 8: Test result.

(MAPE), and R-squared method. The mean squared error (MSE) is shown by

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2. \tag{8}$$

The mean absolute percentage error (MAPE) is shown in the following equation:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|. \tag{9}$$

For R square, we have the estimated value as

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n}y_i. \tag{10}$$

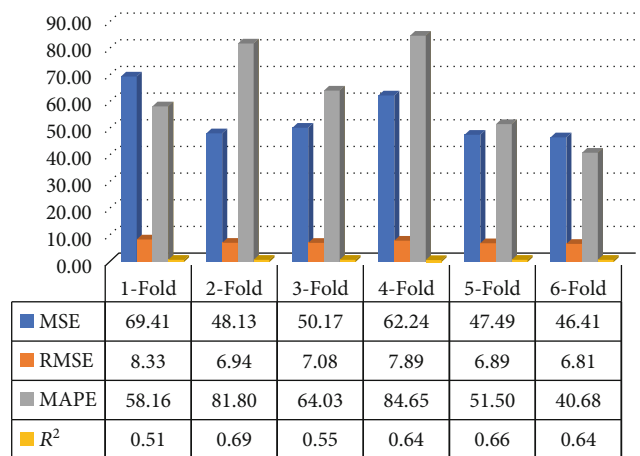|	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	6-Fold
■ MSE	69.41	48.13	50.17	62.24	47.49	46.41
■ RMSE	8.33	6.94	7.08	7.89	6.89	6.81
■ MAPE	58.16	81.80	64.03	84.65	51.50	40.68
■ R^2	0.51	0.69	0.55	0.64	0.66	0.64

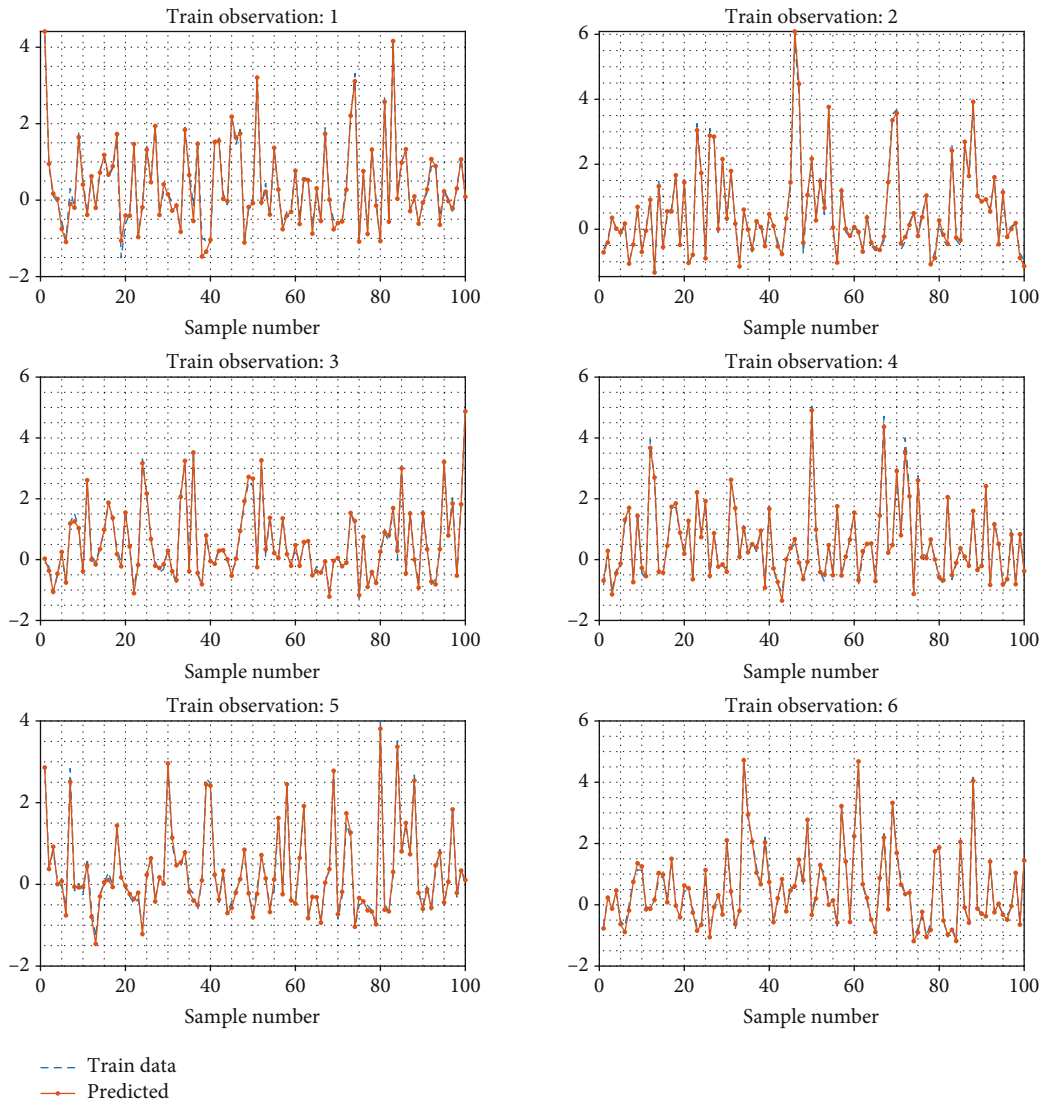FIGURE 9: The comparison of the MSE, RMSE, MAPE, and R for test data.

Figure 10: Train result with PSO.

Then, the variability of the data set can be measured using three sums of squares formulas. The total sum of squares is proportional to the variance of the data:

$$SS_{\text{tot}} = \sum_i (y_i - \bar{y})^2. \tag{11}$$

The regression sum of squares is also called the explained sum of squares:

$$SS_{\text{reg}} = \sum_i (f_i - \bar{y})^2. \tag{12}$$

The sum of squares of residuals is the residual sum of squares:

$$SS_{\text{res}} = \sum_i (y_i - f_i)^2 = \sum_i e_i^2. \tag{13}$$

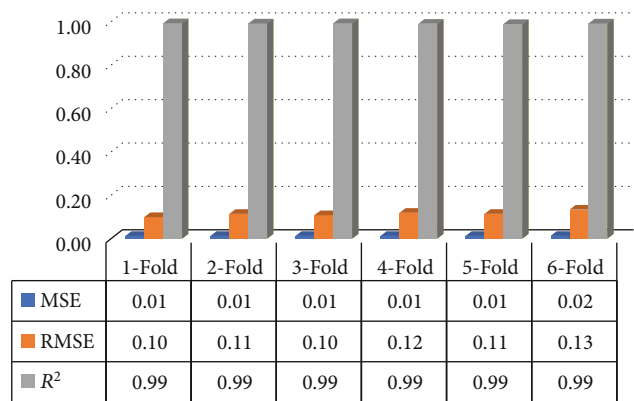|	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	6-Fold
MSE	0.01	0.01	0.01	0.01	0.01	0.02
RMSE	0.10	0.11	0.10	0.12	0.11	0.13
R^2	0.99	0.99	0.99	0.99	0.99	0.99

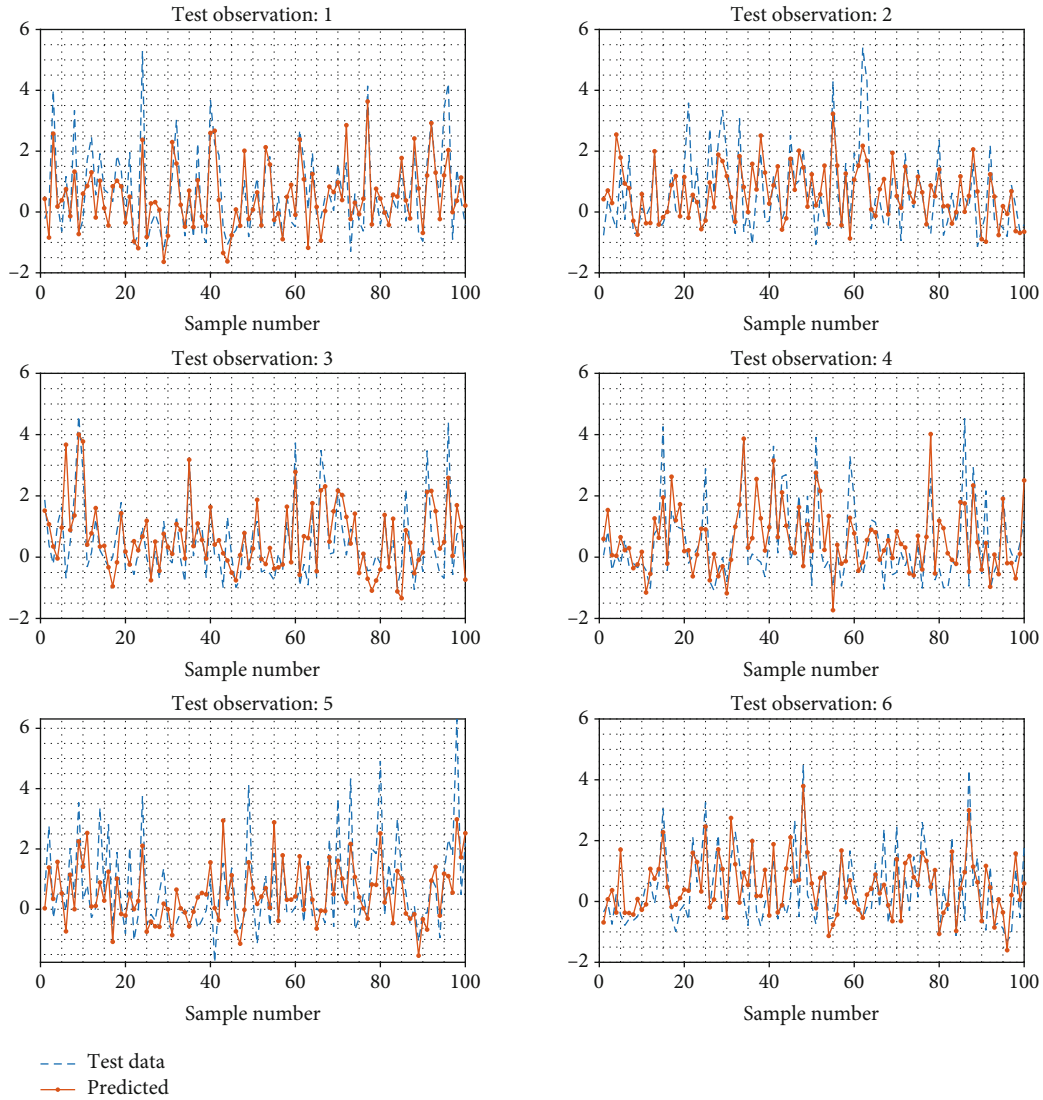Figure 11: The comparison of the MSE, RMSE, and R for train data.
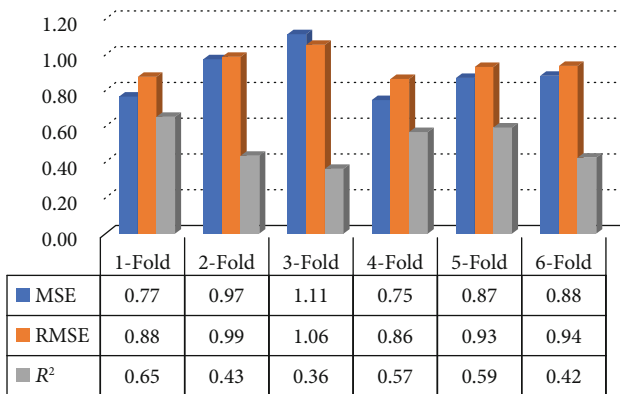
Figure 12: Test result with PSO.

Figure 13: The comparison of the MSE, RMSE, and R for train data.

Table 3: Comparing other methods with the proposed methods.

Method	Recognition rate
PCA [22]	53.2%
SRC [56]	75.12%
CRC [69]	79.4%
Gabor wavelet with Euclidian method [57]	83.44%
Symmetrical face sample method [49]	81.43%
Proposed method	85.25%

The most general definition of the coefficient of determination is

$$R^2 \equiv 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}. \tag{14}$$

Table 2 shows the specifications of the layer that is used in deep learning.

The procedure and test were performed using actual with symmetrical species from ORL and YALE datasets. The results are shown in Figures 6–9. The results show that the system using the ORL dataset revealed how the preprocessing stage improves the accuracy. They also indicate how we can merge or fuse two methods of feature extraction to produce a powerful third method that can accomplish the job.

The comparison of the MSE, RMSE, MAPE, and R for train data is shown in Figure 7.

The result using the PSO is shown in Figures 10–13.

We have observed that the recognition rate and accuracy results from the experiments cannot be met when utilizing the Gabor wavelet and deep learning due to some variation of the values of features which corrupts the classification step. So, when compared with Gabor wavelet features, the variety will be large. Hence, the features are between -14 and 254. Therefore, optimum features are chosen.

PSO methods try to address this problem by selecting only the optimum features from Gabor wavelet. The performance of the classifier is based on the number of features. Too less or too redundant features can reduce the accuracy rates. Therefore, the number of features must be chosen carefully. In PSO, the basic process is that there are a number of particles; each one of them is flying through the problem area arbitrarily searching for the previous best solution and the global best solution of the whole swarm. Then, velocity is modified at each iteration which will define the movement of the particles to be more or less random. Therefore, the algorithms are converged. This method was used in literature [68], using the PSO method for selecting the best features.

In our experiments, we have used Gabor wavelet for feature extraction obtaining 10304 features. When the features were extracted, the implementation of PSO reduces the features to 5142. The best and most optimum features are selected by eliminating the highest and lowest values of features using the fitness function which determines the features that are the closest to each other in the amount. The experimental results obtained a 96% recognition rate on the ORL database when implementing the proposed method.

For the cause of completeness, we compare the performance of PCA [22], SRC [56], CRC [69], Gabor wavelet with Euclidian method [57], symmetrical face sample method [49], and the proposed method.

The comparison of other methods with the proposed methods is shown in Table 3.

5. Conclusion

The use of the symmetry property of the face is an efficient way to increase the performances of the face recognition systems. In this study, a new method is provided for the face recognition system. The new method is upgraded to use the benefits of symmetry property in the face data. The feature space is another way to implement the use of symmetry property in the face. There are many methods for feature extraction; however, none of them can handle the symmetry procedure in the feature space. The suggested methods can perform the symmetry procedure either in the image space or in the feature space. The introduced method is examined and tested for face recognition using data from ORL and YALE datasets.

Data Availability

All data available for readers are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] J. Bouguila and H. Khochtali, "Facial plastic surgery and face recognition algorithms: interaction and challenges. A scoping review and future directions," *Journal of Stomatology, Oral and Maxillofacial Surgery*, vol. 121, no. 6, pp. 696–703, 2020.

[2] W. W. Bledsoe, "Semiautomatic facial recognition," Technical Report SRI Project 6693, 1968.

[3] M. Yazdi, S. Mardani-Samani, M. Bordbar, and R. Mobaraki, "Age classification based on RBF neural network," *Canadian Journal on Image Processing and Computer Vision*, vol. 3, no. 2, pp. 38–42, 2012.

[4] W.-B. Horng, C.-P. Lee, and C.-W. Chen, "Classification of age groups based on facial features," *Journal of Applied Science and Engineering*, vol. 4, no. 3, pp. 183–192, 2001.

[5] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," *European conference on computer vision*, pp. 469–481, 2004.

[6] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 915–928, 2007.

[7] M. Chandra Mohan, V. Vijaya Kumar, and A. Damodaram, "Novel Method of Adulthood Classification Based on Geometrical Features of Face," *GVIP Journal of Graphics, Vision and Image Processing*, 2010.

[8] V. V. Kumar, G. S. Murty, and P. S. Kumar, "Classification of facial expressions based on transitions derived from third order neighborhood LBP," *Global Journal of Computer Science and Technology*, vol. 14, no. 1-F, 2014.

[9] K. L. Kroeker, "Face recognition breakthrough," *Communications of the ACM*, vol. 52, no. 8, pp. 18-19, 2009.

[10] Ming Zhang and J. Fulcher, "Face recognition using artificial neural network group-based adaptive tolerance (GAT) trees," *IEEE Transactions on Neural Networks*, vol. 7, no. 3, pp. 555–567, 1996.

[11] X. Feng, M. Pietikainen, and A. Hadid, "Facial expression recognition with local binary patterns and linear programming," *Pattern Recognition And Image Analysis C/C of Raspoznavaniye Obrazov I Analiz Izobrazhenii*, vol. 15, no. 2, p. 546, 2005.

[12] M. Elad, R. Goldenberg, and R. Kimmel, "Low bit-rate compression of facial images," *IEEE Transactions on Image Processing*, vol. 16, no. 9, pp. 2379–2383, 2007.

[13] A. Skodras, C. Christopoulos, and T. Ebrahimi, "The jpeg 2000 still image compression standard," *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 36–58, 2001.

[14] S. Rakshit and D. M. Monro, "An evaluation of image sampling and compression for human iris recognition," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 3, pp. 605–612, 2007.

[15] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Face recognition using LDA-based algorithms," *IEEE Transactions on Neural Networks*, vol. 14, no. 1, pp. 195–200, 2003.

[16] Z. Lu and Linghua Zhang, "Face recognition algorithm based on discriminative dictionary learning and sparse representation," *Neurocomputing*, vol. 174, pp. 749–755, 2016.

[17] J. Chen, V. M. Patel, L. Liu et al., "Robust local features for remote face recognition," *Image and Vision Computing*, vol. 64, pp. 34–46, 2017.

[18] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: a literature survey," *ACM Computing Surveys (CSUR)*, vol. 35, no. 4, pp. 399–458, 2003.

[19] L. Wiskott, J.-M. Fellous, N. Kuiger, and C. Von Der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775–779, 1997.

[20] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: illumination cone models for face recognition under variable lighting and pose," *Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643–660, 2001.

[21] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[22] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proceedings 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 586–591, Maui, HI, USA, 1991.

[23] M. Yang and L. Zhang, "Gabor feature based sparse representation for face recognition with Gabor occlusion dictionary," in *Computer Vision–ECCV 2010*, Lecture Notes in Computer Science, pp. 448–461, 2010.

[24] G. Guo and N. Zhang, "A survey on deep learning based face recognition," *Computer Vision and Image Understanding*, vol. 189, article 102805, 2019.

[25] F. V. Massoli, G. Amato, and F. Falchi, "Cross-resolution learning for face recognition," *Image and Vision Computing*, vol. 99, article 103927, 2020.

[26] M. Iqbal, M. S. I. Sameem, N. Naqvi, S. Kanwal, and Z. Ye, "A deep learning approach for face recognition based on angularly discriminative features," *Pattern Recognition Letters*, vol. 128, pp. 414–419, 2019.

[27] C. Xu, Q. Liu, and M. Ye, "Age invariant face recognition and retrieval by coupled auto-encoder networks," *Neurocomputing*, vol. 222, pp. 62–71, 2017.

[28] C.-K. Tran, C.-D. Tseng, and T.-F. Lee, "Improving the face recognition accuracy under varying illumination conditions for local binary patterns and local ternary patterns based on weber-face and singular value decomposition," in *2016 3rd International Conference on Green Technology and Sustainable Development (GTSD)*, pp. 5–9, Kaohsiung, Taiwan, November 2016.

[29] S. Nikan and M. Ahmadi, "A modified technique for face recognition under degraded conditions," *Journal of Visual Communication and Image Representation*, vol. 55, pp. 742–755, 2018.

[30] T.-K. Kim and J. Kittler, "Locally linear discriminant analysis for multimodally distributed classes for face recognition with a single model image," *Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 318–327, 2005.

[31] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face recognition using Laplacianfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 328–340, 2005, http://ieeexplore.ieee.org/ielx5/34/30209/01388260.pdf?tp=&arnumber=1388260&isnumber=30209.

[32] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," in *1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 84–91, Seattle, WA, USA, June 1994.

[33] R. Gross, I. Matthews, and S. Baker, "Eigen light-fields and face recognition across pose," in *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition*, pp. 1–7, Washington, DC, USA, May 2002.

[34] C. Zhao, X. Li, and Y. Dong, "Learning blur invariant binary descriptor for face recognition," *Neurocomputing*, vol. 404, pp. 34–40, 2020.

[35] J. Mairal, J. Ponce, G. Sapiro, A. Zisserman, and F. R. Bach, "Supervised dictionary learning," *Advances in Neural Information Processing Systems*, pp. 1033–1040, 2009.

[36] M. A. Sunday, P. A. Patel, M. D. Dodd, and I. Gauthier, "Gender and hometown population density interact to predict face recognition ability," *Vision Research*, vol. 163, pp. 14–23, 2019.

[37] S. Inamizu, E. Yamada, K. Ogata, T. Uehara, J.-i. Kira, and S. Tobimatsu, "Neuromagnetic correlates of hemispheric specialization for face and word recognition," *Neuroscience Research*, vol. 156, pp. 108–116, 2019.

[38] S. K. Zhou and R. Chellappa, "Illuminating light field: image-based face recognition across illuminations and poses," in *Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings*, pp. 229–234, Seoul, Korea (South), May 2004.

[39] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1063–1074, 2003.

[40] L. Zhang and D. Samaras, "Face recognition from a single training image under arbitrary unknown lighting using spherical harmonics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 3, pp. 351–363, 2006.

[41] V. Blanz, K. Scherbaum, T. Vetter, and H. P. Seidel, "Exchanging faces in images," in *Computer Graphics Forum*, vol. 23, no. 3pp. 669–676, Wiley Online Library, 2004.

[42] J. Royer, C. Blais, I. Charbonneau et al., "Greater reliance on the eye region predicts better face recognition ability," *Cognition*, vol. 181, pp. 12–20, 2018.

[43] M. Kas, Y. el merabet, Y. Ruichek, and R. Messoussi, "Mixed neighborhood topology cross decoded patterns for image-based face recognition," *Expert Systems with Applications*, vol. 114, pp. 119–142, 2018.

[44] A. Shashua and T. Riklin-Raviv, "The quotient image: class-based re-rendering and recognition with varying illuminations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 129–139, 2001.

[45] S. Zhou and R. Chellappa, "Rank constrained recognition under unknown illuminations," in *IEEE International Workshop on Analysis and Modeling of Faces and Gestures, 2003. AMFG 2003*, pp. 11–18, Nice, France, October 2003.

[46] R. Basri and D. W. Jacobs, "Lambertian reflectance and linear subspaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 2, pp. 218–233, 2003.

[47] R. Ramamoorthi, "Analytic PCA construction for theoretical analysis of lighting variability in images of a Lambertian

object," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 10, pp. 1322–1333, 2002.

[48] Y. Gao and H. J. Lee, "Cross-pose face recognition based on multiple virtual views and alignment error," *Pattern Recognition Letters*, vol. 65, pp. 170–176, 2015.

[49] Z. Liu, J. Pu, Q. Wu, and X. Zhao, "Using the original and symmetrical face training samples to perform collaborative representation for face recognition," *Optik-International Journal for Light and Electron Optics*, vol. 127, no. 4, pp. 1900–1904, 2016.

[50] M. A. O. Vasilescu and D. Terzopoulos, "Multilinear subspace analysis of image ensembles," in *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings*, Madison, WI, USA, June 2003.

[51] R. Wang, S. Shan, X. Chen, and W. Gao, "Manifold-manifold distance with application to face recognition based on image set," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, Anchorage, AK, USA, June 2008.

[52] J. B. Tenenbaum and W. T. Freeman, "Separating style and content with bilinear models," *Neural Computation*, vol. 12, no. 6, pp. 1247–1283, 2000.

[53] D. Shin, H.-S. Lee, and D. Kim, "Illumination-robust face recognition using ridge regressive bilinear models," *Pattern Recognition Letters*, vol. 29, no. 1, pp. 49–58, 2008.

[54] S. J. Prince, J. Warrell, J. H. Elder, and F. M. Felisberti, "Tied factor analysis for face recognition across large pose differences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 6, pp. 970–984, 2008.

[55] A. Elgammal and C.-S. Lee, "Separating style and content on a nonlinear manifold," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004*, pp. I-478–I-485, Washington, DC, USA, July 2004.

[56] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Yi Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[57] S. Allagwail, O. S. Gedik, and J. Rahebi, "Face recognition with symmetrical face training samples based on local binary patterns and the Gabor filter," *Symmetry*, vol. 11, no. 2, p. 157, 2019.

[58] J.-K. Kamarainen, V. Kyrki, and H. Kalviainen, "Invariance properties of Gabor filter-based features-overview and applications," *IEEE Transactions on Image Processing*, vol. 15, no. 5, pp. 1088–1099, 2006.

[59] S. Meshgini, A. Aghagolzadeh, and H. Seyedarabi, "Face recognition using Gabor-based direct linear discriminant analysis and support vector machine," *Computers & Electrical Engineering*, vol. 39, no. 3, pp. 727–745, 2013.

[60] M. Haghighat, S. Zonouz, and M. Abdel-Mottaleb, "Identification using encrypted biometrics," in *International Conference on Computer Analysis of Images and Patterns*, Lecture Notes in Computer Science, pp. 440–448, Springer, 2013.

[61] A. Banks, J. Vincent, and C. Anyakoha, "A review of particle swarm optimization. Part II: hybridisation, combinatorial, multicriteria and constrained optimization, and indicative applications," *Natural Computing*, vol. 7, no. 1, pp. 109–124, 2008.

[62] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95-International Conference on Neural Networks*, pp. 1942–1948, Perth, WA, Australia, December 1995.

[63] F. Hafiz, A. Swain, N. Patel, and C. Naik, "A two-dimensional (2-D) learning framework for particle swarm based feature selection," *Pattern Recognition*, vol. 76, pp. 416–433, 2018.

[64] J. Kennedy and R. C. Eberhart, "A discrete binary version of the particle swarm algorithm," in *1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation*, pp. 4104–4108, Orlando, FL, USA, October 1997.

[65] Y. Shi, "Particle swarm optimization: developments, applications and resources," in *Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat. No. 01TH8546)*, pp. 81–86, Seoul, Korea (South), May 2001.

[66] Y. Shi and R. Eberhart, "A modified particle swarm optimizer," in *1998 IEEE International Conference on Evolutionary Computation Proceedings. IEEE World Congress on Computational Intelligence (Cat. No. 98TH8360)*, pp. 69–73, Anchorage, AK, USA, May 1998.

[67] A. Unler and A. Murat, "A discrete particle swarm optimization method for feature selection in binary classification problems," *European Journal of Operational Research*, vol. 206, no. 3, pp. 528–539, 2010.

[68] J. Too, A. R. Abdullah, N. Mohd Saad, and W. Tee, "EMG feature selection and classification using a Pbest-guide binary particle swarm optimization," *Computation*, vol. 7, no. 1, p. 12, 2019.

[69] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: which helps face recognition?," in *2011 International Conference on Computer Vision*, pp. 471–478, NW Washington, DCUnited States, 2011.