IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 8, 2021, accepted March 21, 2021, date of publication March 30, 2021, date of current version April 12, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3069881

Shape and Texture Aware Facial Expression Recognition Using Spatial Pyramid Zernike Moments and Law's Textures Feature Set

VIJAYALAKSHMI G. V. MAHESH[1], CHENGJI CHEN[2], VIJAYARAJAN RAJANGAM[3], (Member, IEEE), ALEX NOEL JOSEPH RAJ[2], (Member, IEEE), AND PALANI THANARAJ KRISHNAN[4], (Member, IEEE)

[1]Department of Electronics and Communication, BMS Institute of Technology and Management, Bengaluru 560064, India
[2]Department of Electronic Engineering, College of Engineering, Shantou University, Shantou 515063, China
[3]Centre for Healthcare Advancement, Innovation and Research, Vellore Institute of Technology, Chennai 600127, India
[4]Department of Electronics and Instrumentation Engineering, St. Joseph's College of Engineering, Chennai 600119, India

Corresponding author: Alex Noel Joseph Raj (jalexnoel@stu.edu.cn)

This work was supported by the Scientific Research Grant of Shantou University, China, under Grant NTF17016.

ABSTRACT Facial expression recognition (FER) requires better descriptors to represent the face patterns as the facial region changes due to the movement of the face muscles during an expression. In this paper, a method of concatenating spatial pyramid Zernike moments based shape features and Law's texture features is proposed to uniquely capture the macro and micro details of each facial expression. The proposed method employs multilayer perceptron and radial basis function feed forward artificial neural networks for recognizing the facial expressions. The suitability of the features in recognizing the expressions is explored across the datasets independent of the subjects or persons. The experiments conducted on JAFFE and KDEF datasets demonstrate that the concatenated feature vectors are capable of representing the facial expressions with better accuracy and least errors. The radial basis function based classifier delivers a performance with an average recognition accuracy of 95.86% and 88.87% on the JAFFE and KDEF datasets respectively for subject dependent FER.

INDEX TERMS Facial expressions, emotions, Zernike moments, Law's texture features, neural network classifier.

I. INTRODUCTION

A face is a unique trait for identifying or recognizing people around us. Facial expressions are the reflection of emotional states and play an important role in non-verbal communication. A facial expression is a result of the motions of the muscles underneath the skin of the face. Analysis of these expressions helps in understanding the behavior of a person and certain anatomical changes. For example, the heart rate is higher in anger as compared to happiness and the skin resistance decreases during sadness revealing high stress. Thus, automated FER and subsequent analysis have found its applications in various domains such as surveillance, crowd emotion monitoring, psychological disorder detection, human-computer interaction, driver safety assistance and so on. Ekman and Friesen [1] has formalized six universal facial expressions such as surprise, fear, disgust, contempt, anger, sadness and happiness. These expressions have evolved through social learning and very much essential for survival.

FER is a pattern recognition system and the basis for FER is observation/identification, understanding and interpretation of the visual cues on the face. The main components of FER are (i) face detection (ii) feature extraction and (iii) expression classification. The spatial arrangement of the facial features like shape, fine lines, wrinkles on the facial skin and the structural elements of the face such as forehead, eyebrow, eyes, mouth create different patterns on the face during an expression. These patterns form the key characteristics comprising both micro and macro details. The micro details include variation in the wrinkles and fine lines changing the facial appearance and texture of the skin. Whereas, the movements in the structural elements of the face indicates the macro details. A raised and arched eyebrow

The associate editor coordinating the review of this manuscript and approving it for publication was Rosalia Maglietta.

with eyes wide open represents surprise. A lowered eyebrow with intensely staring eyes indicates anger. Disgust expression is expressed with a wrinkled nose and lowered eyebrows. Similarly, an open mouth represents fear whereas corners raised mouth shows happiness. The extraction of these patterns forms the feature descriptors for recognizing facial expressions.

Feature extraction and description play a major role in deciding the accuracy of the FER system. The features should be redundant, reliable, robust and unique with the best discriminating ability. Extracting the right features is critical and also a challenging task. Most of the FER methods have been proposed for capturing the facial expressions through shape and texture features as descriptors that include Zernike moments (ZM), histogram of oriented gradients (HOG), active shape model, local binary pattern (LBP), local directional pattern (LDP), statistical measures and gray level co-occurrence matrix (GLCM) [2]. These methods extract either the macro or micro details alone to describe the facial expressions. But, it is important to provide a maximum possible description to enhance the performance of the system. Thus, both texture and shape information provided by the Laws texture energy measures and Zernike Moments are utilized for FER. Texture information is obtained by five types of kernels such as level, edge, spot, ripple and wave. Distinct information is obtained with each kernel revealing the changes in the wrinkles and fine lines of the facial skin. Zernike moments provide the shape information of the changes in the facial appearance due to the movements in the structural elements. In this perspective, ZM is extended to spatial pyramid representation with three-level decomposition to capture the shape information in each facial sub-region thus forming Spatial pyramid Zernike moments (SPZM). The Law's texture energy features (LTexM) and SPZM are combined to form the integrated feature vector that has improved discriminating ability to recognize and classify the facial expressions. The contributions of this paper are (i) an integrated feature set for capturing the macro and micro details from the facial expressions using SPZM and Law's texture features (ii) quantitative assessment of improved recognition accuracy by considering the images with different orientations under subject dependent and independent FER (iii) integration of features for the effective representation of the facial expressions (iv) robust features for recognizing the facial expressions from the full left and right profile of facial images.

The literature review is presented in section II. Section III describes the Law's texture energy measures and SPZM. The proposed methodology and the experimental results are presented in sections III and IV respectively. Conclusion is presented in section V

II. RELATED WORK

The different approaches for extracting the facial feature to recognize the expressions include (i) Facial action coding system (FACS) and Action unit (AU) (ii) geometric, appearance and hybrid methods. In AU method, the movement of muscles responsible for producing a facial expression is encoded into 46 facial AUs [3]. FER system detects the face AUs to classify the expressions using observations and comparisons. FACS describes the relationship between the muscle movements of the face and expressions and was introduced by Ekman and Friesen [1] based on the characteristics of AUs. The second method depends on the extraction of content-based facial features. It depends upon the appearance, geometric and hybrid characteristics. Appearance-based approaches are holistic capturing global information from the facial images to generate the feature vector for distinguishing facial expressions. Non-holistic approaches make use of geometric features such as eyes, nose, mouth, chin, head outline of a face and their relationships.

Holistic approaches apply transformations and use statistical methods to extract the features representing the texture characteristics of the image. Gabor filter [4]–[8] provides texture descriptors with good discrimination ability. It provides both magnitude and phase components, but magnitudes are selected as features as they are invariant to translation. Gabor filter provides multidimensional feature vectors with high computation cost and the dimension of features can be reduced by principal component analysis (PCA). LBP [9], [10] generates binary patterns to represent the texture by comparing the center pixel with the neighborhood pixels of a region from the facial image. Variants of LBP were further introduced for improvement in the performance of the FER system. A weighted multi-scale method for LBP was proposed by [11], in which multiple weighted LBP features are extracted with different scales. The extended LBP is combined with the Karhunen-Loeve (KL) transform in [12]. The role of subspace analysis methods such as PCA and Independent Component Analysis (ICA) is investigated for the extraction of the facial expression features [13]. Other methods like local directional numbers (LDN) [14], local ternary pattern [15], discrete wavelet transform (DWT) [16] and sparse local Fisher discriminant analysis (SLFDA) [17] were also explored for FER in recent years.

Non-holistic approaches extract the geometrical features providing the position of facial landmarks and shape information. Geometric features were extracted using the curvelet transform. The coefficients of the transform with varying scales and angles form the feature vectors to represent the facial expressions [18]. Histogram of oriented gradients, originally proposed for object recognition, was found significant for FER. HOG provides the magnitude and phase of the gradients from which the dominant gradients are selected relating to the edge information [19], [20]. In [21], facial landmarks are tracked based on elastic bunch graph matching (EBGM) displacement. The facial landmarks or combination of the landmarks form the feature vector representing a facial expression. In [22], optical flow based facial points were tracked from consecutive frames to detect the movement of facial points to provide dynamic features. Moments based shape descriptors are critical in representing the facial

expressions. Zernike moments with magnitude features, exhibiting orthogonal and rotation invariant properties, are used in [2], [23], [24]. Pseudo Zernike moments (PZM) [25] also provide a good description for FER. Appearance-based methods are affected by lighting and orientation conditions. Geometry based methods provide better feature descriptors working well irrespective of the variations in the facial image.

Further, the recent years have also witnessed the success of deep learning methods with multilayered architectures in facial expression recognition. The deep learning methods automatically compute the features for data representation while reducing the requirement for extracting the hand crafted features. The work presented in [27] utilized the ZM for deriving the coefficients of the convolution kernels in convolution layer of CNN architecture. This was found to be significant in extracting the shape features and improved the classification accuracy. Deep sparse auto encoders were implemented by [28] to learn discriminative and robust features. The work presented in [29] introduced a part based hierarchical bi directional recurrent neural network to analyze the dynamic evolution and morphological variations in facial expressions which proved to be effective in reducing the error rates. Modifications in CNN architectures are introduced by [30]–[33] to enhance the performance of FER system. Generative adversial networks(GAN) [34], [35] with generator and discriminator networks have also emerged to be better models in discriminating the facial expressions. Though deep features are efficient and have outperformed the existing hand crafted feature methods, the deep learning methods require larger datasets for training and are computationally expensive.

Thus this paper proposes a method considering the advantages of appearance and geometric based methods. Hence, the features from holistic and non-holistic approaches are combined to form a robust feature vector for improved classification. The efficacy of the selected feature descriptors in the proposed method is proved by focusing on both subject dependent and subject independent FER.

III. SPZM AND LTexM INTEGRATED FEATURE SET MODEL FOR FER

The schematic of the proposed feature concatenation strategy for recognizing facial expressions is shown in Fig. 1. The FER is a pattern recognition problem where a facial pattern (FP) is assigned one of the m expression labels $\{c_i\}_{i=1,2.....m}$. The key characteristics or patterns are extracted from the facial images with expressions in the form of feature vectors. The selection of the right features improves the degree of accuracy in classification. LTex and SPZM are extracted from the facial images and integrated to form a feature vector for training a neural network classifier. Upon training, the classifier model is observed to classify the facial expression of an image. Finally, the performance of the model is quantitatively assessed using various performance measures based on the elements of the confusion matrix.

The process of the proposed approach is presented in Algorithm 1.

Algorithm 1 The process flow of SPZM and LTexM integrated feature set model

Begin

Input: Facial expression images database with seven emotions

Dataset division: Training and Testing dataset

Training phase:

1) Assign M=Number of training images
2) **for** i=1 to M
   a) Extract the Law's texture features
   b) Extract Zernike moment based shape features from spatial pyramid of the image SPZM = $\{|A_{nm}|_k^l\}$
   c) Integrate the texture and shape features to form integrated feature vector FV_{int} = {LTexM, SPZM}

   end
3) Assign the class labels to the integrated feature set with seven expressions C={angry, disgust, fear, happy, sad, surprise and neutral} and create label feature set matrix $\{FV_{int}, C\}$
4) Provide $\{FV_{int}, C\}$ to the neural network classifiers (MLPNN and RBFNN)
   a) Define the learning parameters of the neural network
   b) Tune the parameters to improve the performance and derive the classifier model

Testing phase (Output):

1) Assign N=Number of testing images
2) **for** i=1 to N
   a) Extract the Law's texture features
   b) Extract Zernike moment based shape features from spatial pyramid of the image SPZM = $\{|A_{nm}|_k^l\}$
   c) Integrate the texture and shape features to form integrated feature vector FV_{intq} = $\{\text{LTexM, SPZM}\}_Q$

   end
3) Provide FV_{intq} to the classifer model to predict the class label which belongs to set C

End

The subsections present the comprehension of the proposed work.

A. FACIAL EXPRESSION DATASET

The two databases (i) Japanese female facial expression (JAFFE) and (ii) Karolina Directed Emotional Faces (KDEF) are experimented to analyze the performance of the proposed FER

The JAFFE database [36] is widely used for evaluating the performance of FER systems. JAFFE is a dataset collected from 10 Japanese females. Each individual has posed with six basic facial emotions like angry, disgust, fear, happy, sad and

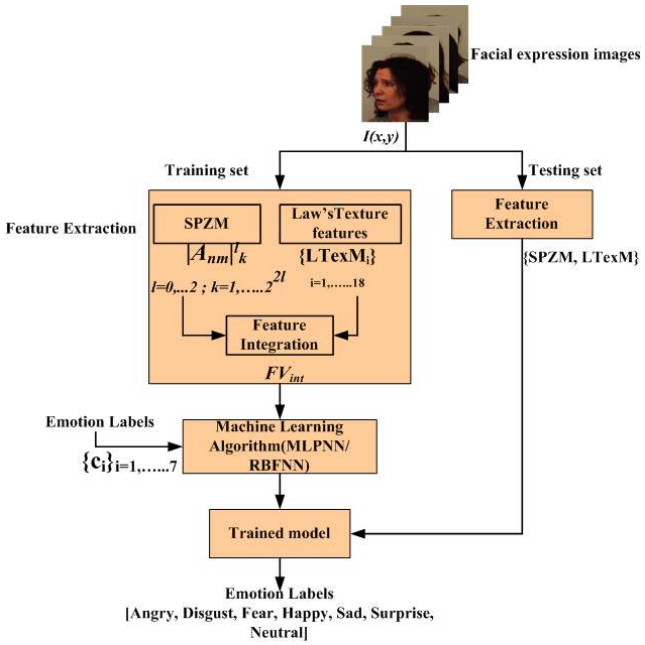

FIGURE 1. Proposed methodology for facial expression recognition.

FIGURE 2. Sample images from the JAFFE database.

surprise with a neutral pose. This dataset has 213 grayscale images with labels of facial expression and each image has a size of 256 × 256 pixels. Fig. 2 shows the sample images of a Japanese female with facial expressions from the database.

The KDEF database [37] developed by the department of clinical neuroscience, a section of psychology from Sweden for psychological and medical research purposes. The database has 4900 color images of facial expressions acquired from 70 individuals (35 females and 35 males) of age ranging from 20 to 30 years. Here, each individual has displayed seven emotions such as angry, disgust, fear, happy, sad, surprise and neutral. The facial expressions of each person is captured under 2 sessions from 5 different angles: -90, -45, 0, 45 and 90 degrees, providing full left, half left, straight, half right and full right profiles under uniform lighting. The spatial resolution of the facial image is 562 × 762. Fig. 3 shows the sample images of an individual with facial expressions from the KDEF database.

For the proposed work, 700 facial expression images are acquired from 20 female subjects with all five orientations. All the images were converted to gray-scale for computational reasons. The problem of facial expression classification is more challenging with this dataset as some images are captured with different orientations and only partial face is visible in full left and full right profiles. Thus, the FER system with a robust feature extraction method is required to overcome the complexity of the problem.

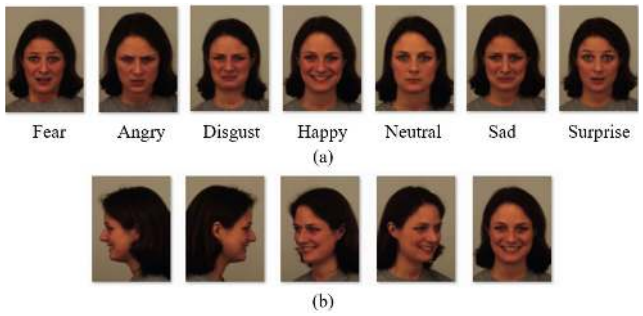

FIGURE 3. Sample images (a) the emotions of an individual (KDEF database) and (b) five different profiles/orientations of an expression.

The performance of the proposed framework was validated considering subject dependent and subject independent facial expression recognition.

1) SUBJECT DEPENDENT FER

For subject dependent expression recognition, the JAFFE and KDEF databases were used individually. A single dataset, containing the facial expressions of all the subjects, is divided into training and testing sets with no overlapping using hold out(HO) method. 5% of the images from the JAFFE database and 60% from KDEF database are used for training to build the classifier model. The remaining samples are used to evaluate the model after testing. Though the testing and training sets are disjoint, they are framed considering the images of all the subjects from the dataset as shown in Fig. 4(a) for the JAFFE database.

2) SUBJECT INDEPENDENT FER

For Subject independent FER, the subjects or persons considered with their facial expressions for training and testing should be disjoint. Thus the images of the 20 female subjects from the KDEF database were considered for training while the facial expression images of the 10 subjects from the JAFFE database were used for testing the model.

B. FEATURE EXTRACTION

Features were extracted from the facial expression images to carry out classification and recognition.

1) FEATURE EXTRACTION USING ZERNIKE MOMENTS

During a facial expression, a lot of movements in the structural elements of the face: eyebrows, eyes, nose, mouth and chin can be observed, thus resulting in changes in the appearance of the face. The most prominent change is in the shape characteristics of the face. These changes need to be captured and represented by a suitable shape descriptor to recognize and classify the facial expression to interpret the emotion. The preferred shape descriptor should have two characteristics to provide the best discrimination between the facial expressions. The characteristics are (i) invariance to rotation, translation and scaling (ii) lower redundancy between the features. Zernike Moments [38], from the class

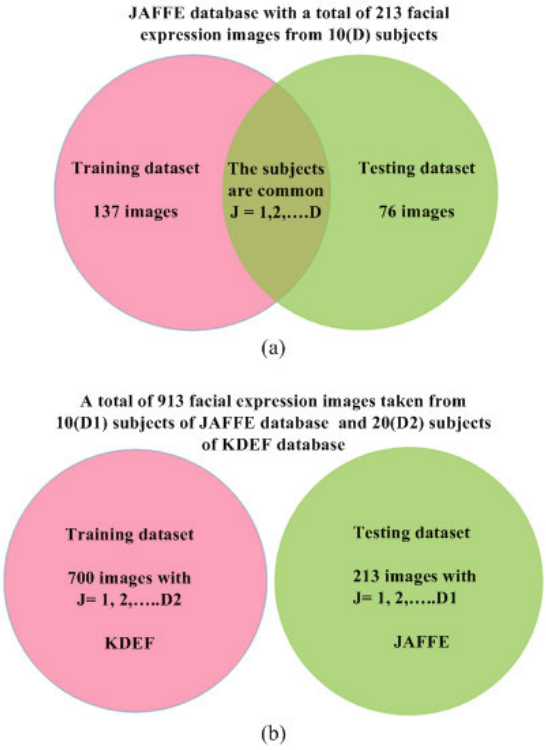

FIGURE 4. Dataset division (a) Subject dependent-JAFFE and (b) Subject independent FER.

of orthogonal moments, proved acceptable to be the shape descriptors with a higher degree of information satisfying the above requirements [39]–[41]. The orthogonality of the moments provides lower possible redundancy and the magnitude of the ZM is invariant to rotation. Further, the lower and higher-order moments provide global shape and detailed information respectively. Thus, to capture the complete shape information from an image, a set of ZM is computed by varying the moment orders from a lower value to the higher. This process involves a significantly larger number of computations.

To reduce the complexity of the computation, the size of the set of ZM is predefined and combined with the spatial pyramid of the image to obtain the spatial pyramid ZM feature. A key characteristic of ZM that best complies with the spatial pyramid of the image is its hierarchical nature. The lower order moments provide the global shape information and the higher-order moments reveal the local shape information.

2) SPZM FEATURES

The SPZM feature is motivated by the concept of spatial pyramid representation of an image [42], [43], the shape descriptors and, Zernike moments. The idea is to extract the shape information at different levels of the pyramid to create a multilevel shape representation to capture the complete pattern created from the facial expression. The ZMs are computed at each pyramid level and are concatenated to form SPZM feature vector. Thus, this descriptor can be viewed as

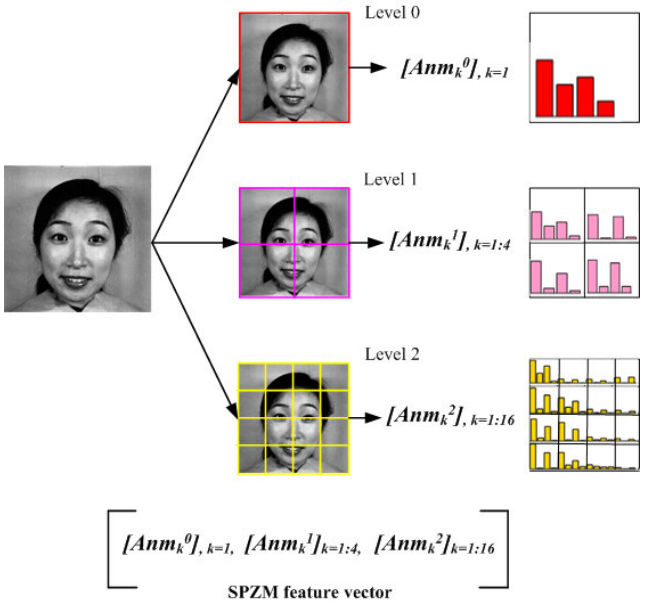

FIGURE 5. SPZM feature of a sample image.

a combination of global shape characteristics and local shape characteristics of an image that can efficiently discriminate facial expressions. The extraction of SPZM involves (i) construction of a set of grids at different levels. (ii) computation of ZM for each sub-region or cell of the pyramid levels and concatenation of the moments.

The facial image $I(x, y)$ is divided into a set of grids at levels 0,..,L where L=2, such that the grid at each level l has 2^{2l} sub-regions or cells $C(x, y) = I(x, y)_k^l$ along each axis with $l = 0,\ldots\ldots\text{L}$ and $k = 1, \ldots. 2^{2l}$. Now for each cell $C(x, y)$, ZM is computed to extract the features. The ZM is computed for all the cells at different levels of the image pyramid and integrate them to form SPZM feature as shown in Fig. 5.

The ZM is computed by projecting each cell $C(x, y)$ on to the set of complex Zernike polynomials $V_{nm}(x, y)$

$$A_{nm} = \frac{m+1}{\pi} \sum_x \sum_y C(x, y) V_{nm}^*(\rho, \theta) \quad with\ x^2 + y^2 \leqslant 1 \tag{1}$$

such that

$$V_{nm}(x, y) = V_{nm}(\rho, \theta) = R_{nm}(\rho) e^{jm\theta} \tag{2}$$

where,

n is the order of the polynomial

m is the repetition factor such that $|m| \leqslant n$ and $n - |m|$ is even

ρ is the length of the vector from the origin to the pixel located at spatial location (x, y) and is presented by $\rho = \sqrt{x^2 + y^2}$

θ angle of the vector from origin to the pixel located at spatial location, (x, y) from the x-axis in counter clockwise direction and

$R_{nm}(\rho)$ is the radial polynomial defined as

$$R_{nm}(\rho)=\sum_{s=0}^{n-|m|}(-1)^s\frac{(2n+1-s)!}{s!(n-|m|-s)!(n+|m|+1-s)!}\rho^{n-s} \tag{3}$$

The ZM thus computed is a complex quantity, from which the magnitude $|A_{nm}|$ and phase Arg (A_{nm}) are obtained. The magnitude $|A_{nm}|$ is selected since it is preserved as the shape descriptor by varying the order of the moment and the repetition factor. For a cell $C(x, y)$ a set of ZM magnitudes $[|A_{11}|, |A_{20}|, |A_{22}|, |A_{31}|]$ is computed by varying the moment order from 1 to 3 with the associated repetition factors. The moment $|A_{00}|$ is ignored as the zeroth-order Zernike polynomial. V_{00} is flat, in turn, an image projected on to this polynomial does not provide any edge maps or shape information. In ZM, the order of the polynomial indicates the radial component and the repetition factor indicates the sinusoidal component. With higher-order, the number of turning points in Zernike polynomial rises and thus provides better shape information of an image. The next higher-order polynomials next to zero are considered for image description in this work. The Zernike polynomial basis functions $V_{00}, V_{11}, V_{20}, V_{22}, V_{31}$ over a unit circle and their magnitudes $|V_{nm}|$ are provided in Fig. 6(a) and Fig. 6(b). Fig. 6(b) illustrates the rotational invariance characteristics of the ZM. The one dimension profile of the 2D Zernike polynomials is displayed in Fig. 6(c) to emphasize the number of turning points obtained with a variation in moment order. The complete shape description for an image (Fig. 5) is obtained by concatenating $[|A_{11}|, |A_{20}|, |A_{22}|, |A_{31}|]$ obtained for each $C(x, y)$ at levels $l = 0,1,2$ as indicated in equation (4) to form SPZM features.

$$\text{SPZM}=\{[|A_{nm}|^0_{k\,k=1}], [|A_{nm}|^1_{k\,k=1:4}], [|A_{nm}|^2_{k\,k=1:16}]\}_{n=1,2,3} \tag{4}$$

3) LAW's TEXTURE FEATURES

Texture features provide the statistical measures of an image based on the spatial arrangement of the pixel intensities. In the proposed work, the texture features are utilized to capture the micro details of the facial expression reflecting the texture of the skin through the formation of wrinkles and fine lines. These details are significant and contribute to identifying and recognizing the emotional state associated with expression. Kenneth I Laws [44] proposed texture energy measures for classifying the textures. These features are invariant to changes in rotation, contrast and luminance. These measures find applications in various domains [45]–[47].

The Law's texture feature extraction method is presented as,

1) Law's texture features used a set of one-dimensional convolution masks that are center-weighted and symmetrical or anti-symmetrical with varying dimensions of 1×3, 1×5 and 1×7. It is proved that masks with dimension 1×5 provide better texture descriptors. Each of the masks is used to analyze the level, edge, spot and ripple characteristics of the image and named appropriately. The details of the 1×5 masks are as shown

L5(level)	[1 4 6 4 1]
E5(Edge)	[− 1 2 0 2 1]
S5(Spot)	[− 1 0 2 0 − 1]
R5(Ripple)	[1 − 4 6 − 4 1]

2) One dimensional masks were combined to generate two-dimensional masks by computing the outer product of each vector as indicated,

	L5	**E5**	**S5**	**R5**
L5	L5L5	L5E5	L5S5	L5R5
E5	E5L5	E5E5	E5S5	E5R5
S5	S5L5	S5E5	S5S5	S5R5
R5	R5L5	R5E5	R5S5	R5R5

This resulted in a set of 16 convolution masks $\{CM_i\}_{i=1,2,.....16}$, each with a dimension of 5×5. The appropriate CM were selected from the convolution masks based on the listed criteria: (i) the mask L5L5 was dropped as it is sensitive to changes in the intensities and the sum of the elements of the mask adds up to a non zero value. (ii) The masks that provide a similar type of information were combined. For example, L5E5 measures vertical edge content while E5L5 presents horizontal edge content, thus the average will provide the total edge information. Accordingly, a set of 9 convolution masks $\{CM_i\}_{i=1,2.....9}$ were finally selected as presented in equation 5.

$$\{CM_i\}=\{L5E5/E5L5, L5S5/S5L5, L5R5/R5L5, E5S5/S5E5, E5R5/R5E5, R5S5/S5L5, E5E5, S5S5, R5R5\} \tag{5}$$

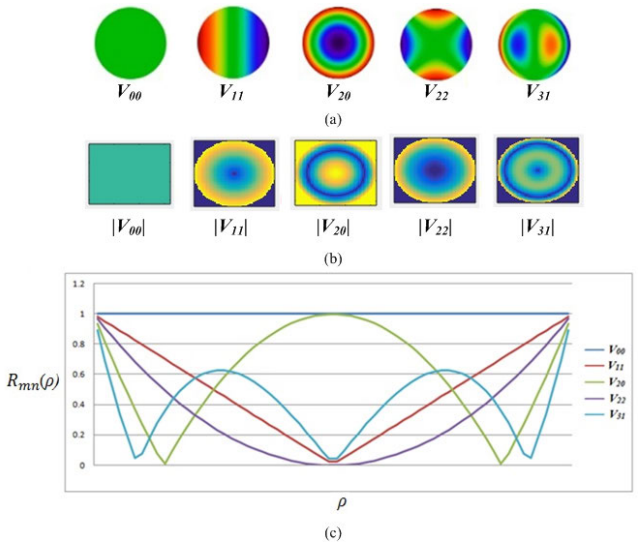

FIGURE 6. Zernike polynomial basis functions (a) $V_{00}, V_{11}, V_{20}, V_{22}, V_{31}$ over a unit circle (b) Magnitudes of the basis functions (c) Zernike polynomials in 1D.

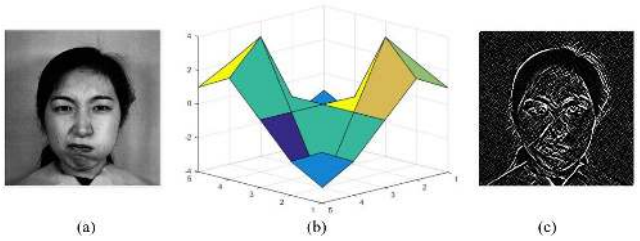

FIGURE 7. Sample of an individual (a) with an angry expression, (b) Convolution mask E5E5 and (c) Law's texture image resulted from convolution with E5E5 mask from the JAFFE database.

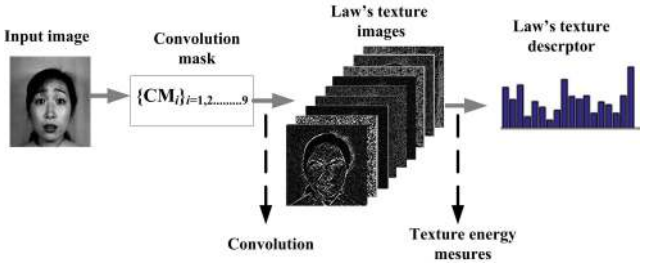

FIGURE 8. Computation of Law's texture feature for a sample image.

3) Later, each of the facial images $I(x, y)$ is convolved with a set of masks $\{CM_i\}_{i=1,2.....9}$ to produce a set of Law's texture images $\{LT_i\}_{i=1,2.....9}$ representing the nine texture descriptors of a pixel for the input image. Fig. 7 presents a sample of the convolution mask with the corresponding texture image.
4) Finally, Law's texture energy measures, mean and standard deviation were computed for every image of the set $\{LT_i\}_{i=1,2.....9}$, which form the texture descriptor with a length of 18 describing the variations in the image during an expression as indicated in Fig. 8.

C. FEATURE INTEGRATION

The Law's texture features and shape features from ZM are concatenated to form an integrated feature vector FV_{int}. As the contributions from the texture and shape features are different, the feature set consolidates the information from both the features using a series fusion rule.

$$FV_{int} = \{\text{Law's texture features, Spatial pyramid ZM features}\} = \{\text{LTexM, SPZM}\} \quad (6)$$

During integration, no weights were considered to give equal preference to both the feature types. The integrated feature set is normalized using Z-score normalization to form a new vector FV_{int_norm} with mean 0 and standard deviation 1. Feature normalization is crucial as it is one of the requirements of the machine learning algorithm and the magnitudes of the features in the feature vector influence the weight update during training process [49]. The normalized feature set is represented by,

$$FV_{int\_norm} = \frac{FV_{int} - \mu}{\sigma} \quad (7)$$

where μ is the mean and σ is the standard deviation of the features. The features are normalized by retaining their original properties.

D. CLASSIFICATION USING ARTIFICIAL NEURAL NETWORKS (ANN)

The extracted texture and shape features from the facial expression image are fused to form an integrated feature set FV_{int_norm}. The concatenated feature vector is provided to the feed forward artificial neural network (ANN) for classification. ANNs are used extensively for facial expression recognition [40] as they have the ability (i) to learn from the examples and model complex non-linear relationships and (ii) to generalize well by predicting the class labels of unseen data. This work proposes to use multilayer perceptron feed forward artificial neural network (MLPNN) and radial basis function neural network for classification

To train both MLPNN and RBFNN, the training images from the JAFFE and KDEF databases were considered as mentioned in section III.A and remaining images were used for testing. The FV_{int_norm} extracted from the training images were labeled for emotions from $C = [c_1, c_2, \ldots, c_7]$, where c_i represents the facial expressions like angry, disgust, fear, happy, sad, surprise and neutral.

1) MULTILAYER PERCEPTRON NEURAL NETWORK

The neural network architecture is designed with an input layer, one hidden layer associated with a hyperbolic tangent activation function $f(x) = (\frac{2}{1+e^{-2x}-1})$ and an output layer with a linear activation function. Each layer is associated with the trainable parameters known as weights and biases. The input layer accepts an input vector, and passes to the hidden layer for processing. The hidden layer output is fed to the output layer to present the network's final output.

The MLPNN network is trained with the labeled feature vector $\{FV_{int\_norm}, \text{C}\}$ to adjust the trainable parameters for tuning the output close to the target values. For training, Levenberg-Marquardt back propagation algorithm [48] was used with mean squared error as the cost function. The stopping criteria was set by defining the number of epochs. MLPNN was tuned by providing performance goal(MSE), learning rate (η), momentum(m) and regularization parameter (λ) to generate the optimized classifier model.

2) RADIAL BASIS FUNCTION NEURAL NETWORK

RBFNN is a three layered network with input, hidden and output layers. The network is robust and converges at a faster rate. RBFNN uses Gaussian activation function from the class of radial functions in the hidden layer which is monotonic with respect to the distance from the center. The Gaussian function with center μ and spread/radius is represented by

$$G(x) = exp(-\frac{(FV_{int\_norm} - \mu)^2}{\sigma^2}) \quad (8)$$

For the known input FV_{int_norm}, the hidden layer computes the distance between the input and centers of the basis functions $\|(FV_{int\_norm} - \mu)\|$, where $\|.\|$ is the L2 norm and applies RBF function. The computation is continued in the output layer that predicts the class label of the sample. The RBFNN is provided with $\{FV_{int\_norm}, C\}$ for training. During training the learnable parameters of the network, spread of the activation function and the weights connecting hidden layer to output layer, are tuned to obtain the best model.

The effectiveness of the models are now tested using the samples from testing set with unseen samples. For the testing sample, the features $FV_{int\_norm\_test}$ are extracted and is fed to classifier model to predict the class label c_{test}. where $c_{test} = f(FV_{int\_norm\_test}; \{FV_{int\_norm}, C\}$, MLPNN/RBFNN parameters) and $c_{test} = \{c_i\}_{i=1,2.....7}$

IV. RESULTS AND DISCUSSION

To evaluate the performance of the proposed method, two experiments (i) subject dependent FER and (ii) subject independent FER were performed on JAFFE and KDEF databases.

A. SUBJECT DEPENDENT FER

The first experiment was conducted to investigate the combination of ZM based shape descriptors and Law's texture features for efficient facial expression representation and discrimination. This was experimented on the JAFFE and KDEF databases separately using the hold out method. As per the method, two disjoint sets were created for training and testing the classifier model.

1) CASE (I)

For all the images of the training set the SPZM and texture features LTexM were computed. The SPZM are extracted by considering the ZM orders $n = 1,2,3$ with the associated repetition factors as presented in equation (4). The SPZM shape features of length 84 are integrated with LTexM and normalized to form $FV_{int\_norm\_1}$ with dimension 1×102. The labeled and normalized feature vector is provided to the neural network classifiers for training the model that can further predict the expression label for the unseen sample.

The MLPNN was trained by providing the labeled training dataset to learn the relationship between the input feature vectors and the class labels by modifying the weights and biases. The network was tuned appropriately by setting the number of epochs limit to 1000 and cost function close to zero. The network converges and provide the best model for $\eta = 0.5$, $m = 0.95$ and $\lambda = 0.6$. The model is tested by providing the facial images from the testing set. The ability of the model in classifying all the facial expressions is assessed by computing the performance metrics from the confusion matrix that include classification accuracy(CA), true negative rate (TNR), false acceptance rate(FAR) and false rejection rate (FRR). Table 1 shows the confusion matrix.

Where, True Positive (TP): true accept.

True Negative (TN): True reject

TABLE 1. Confusion Matrix.

Actual facial expression	Predicted facial expression	
	TP	FN
	FP	TN

TABLE 2. Confusion matrix of MLPNN experimented with JAFFE database.

Angry			Disgust			Fear			Happy	
9	1		5	6		10	2		9	3
1	65		3	62		3	61		3	61

Neutral			Sad			Surprise	
9	1		5	6		10	0
2	64		2	63		0	66

TABLE 3. Performance metrics computed from testing dataset for MLPNN with $FV_{int\_norm\_1}$.

	JAFFE				KDEF			
Emotions	CA(%)	TNR(%)	FAR(%)	FRR(%)	CA(%)	TNR(%)	FAR(%)	FRR(%)
Angry	97.36	98.48	1.51	1.13	77.5	85.8	14.16	8.34
Disgust	88.15	95.38	4.61	7.24	74.28	81.66	18.33	7.39
Fear	93.42	95.31	4.68	1.9	85.71	95.8	4.16	10.13
Happy	92.1	95.31	4.68	3.22	77.5	84.5	15.4	7.1
Neutral	96.05	96.96	3.03	0.92	81.07	91.25	8.75	10.18
Sad	89.47	96.92	3.07	7.46	77.5	88.3	11.6	10.9
Surprise	100	100	0	0	83.57	96.25	6.6	9.83

False Positive (FP): False accept.

False negative (FN): False reject.

Here TP and TN represent the correct facial expression classifications whereas FP and FN are miss classifications, For example, a happy face falsely recognized as sad or angry.

From the result of the testing process, confusion matrices [49] are framed for each expression for both databases as presented in Table 2. From the attributes of the confusion matrix, the performance metrics are calculated.

$$\text{Classification accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

$$\text{TNR/Specificity} = \frac{TN}{TN + FP} \tag{10}$$

$$\text{FAR} = \frac{FP}{TN + FP} \tag{11}$$

$$\text{FRR} = \frac{FN}{FN + TP} \tag{12}$$

The performance metrics computed for all the expressions of the testing dataset from JAFFE and KDEF are shown in Table 3.

The experiment was conducted to prove the discriminating ability of the integrated feature set in the accurate classification of facial expressions. From the results, it is observed that the proposed method delivers more improved classification accuracy. Also, it can be observed that the MLPNN provides high CA (greater than 90%) for the expressions angry, fear, happy, neutral and surprise for the facial images from JAFFE dataset and CA greater than 80% for fear, neutral and surprise from KDEF dataset.

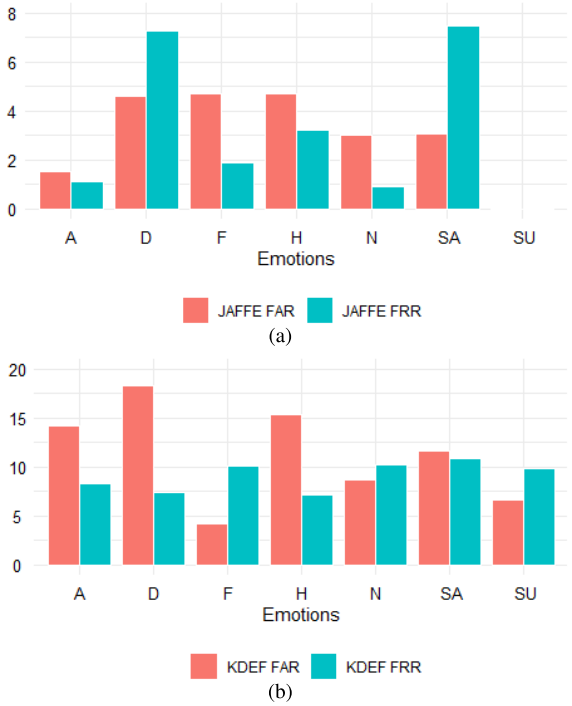

FIGURE 9. Plot of FAR and FRR metrics provided by testing (a) JAFFE and (b) KDEF databases for MLPNN with $FV_{int\_norm\_1}$. A - Angry, D - Disgust, F - Fear, H - Happy, N - Neutral, SA - Sad, SU - Surprise.

Further, the specificity of the system was also measured through TNR. A value, TNR=1(100%), indicates the best specificity. A higher value of TNR close to one produces less false-positive results. From the results of the JAFFE database, it is observed that the average TNR is 96.90%, so the average prediction of false positives is only 3.1%. Similarly, for KDEF dataset, the average TNR is 89%.

In the next experiment, the system was tested for type-I and type-II errors which represent false match and false non-match respectively. FAR indicates type-I error whereas FRR indicates type-II error. Ideally, the two errors are inversely proportional to each other. False acceptance and false rejection recognize the facial expression and assign an emotion class label to which it does not belong to. These two are undesirable and may affect the success rate of a FER system. In some applications like psychological disorder detection, it may lead to false diagnosis and if a device has to take some actions based on human emotions. The plots shown in Fig. 9(a) and Fig. 9(b) display FAR and FRR of both databases. These plots aid in analyzing the relationship between FAR and FRR for each facial expression.

Fig. 9 notifies the inverse relationship between the errors which ensures that the proposed framework returns the least possible errors for identifying the emotions represented by the facial expressions. Thus the proposed system provided an average FAR of 3.08% and FRR 3.12% for JAFFE dataset. Similarly, the two measures are 11.28% and 9.12% respectively for KDEF dataset which is desirable for a FER system to contribute a higher performance rate.

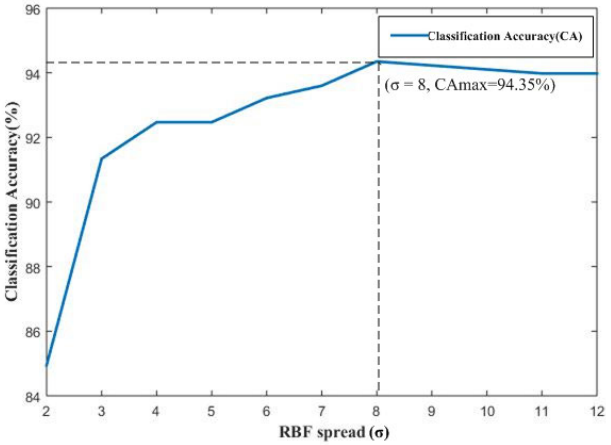

FIGURE 10. Plot of Classification accuracy for variation in RBF spread for $FV_{int\_norm\_1}$.

Later, the RBFNN was trained and tested with the training and testing sets from JAFFE and KDEF datasets. For the training process, the cost function, sum squared error(SSE), was set and the spread(σ) of the radial activation function was varied from $\sigma = 2$ in incremental steps of 1 to improve the performance. With each variation in σ, the accuracy of the model was noted. The network converged for $\sigma = 8$. Fig. 10 shows the accuracy of the model with respect to spread of RBF obtained for JAFFE database.

From the Fig. 10, it is noted that the maximum accuracy of 94.35% is obtained for $\sigma = 8$. With the optimized value, the performance metrics are computed for all the expressions of the testing datasets and are presented in Table 4.

TABLE 4. Performance metrics computed from testing dataset for RBFNN with $FV_{int\_norm\_1}$.

	JAFFE				KDEF			
Emotions	CA(%)	TNR(%)	FAR(%)	FRR(%)	CA(%)	TNR(%)	FAR(%)	FRR(%)
Angry	97.36	98.48	1.51	1.13	74.28	82.08	17.91	7.81
Disgust	89.47	93.84	6.15	4.38	73.92	81.25	18.75	7.33
Fear	90.78	95.31	4.68	4.54	82.14	92.5	7.5	10.36
Happy	96.05	98.43	1.56	2.39	80.35	91.66	8.33	11.32
Neutral	98.68	98.48	1.51	0	79.64	89.58	10.41	9.95
Sad	93.42	96.92	3.07	3.51	83.21	96.25	3.75	13.04
Surprise	94.73	95.45	4.54	0.73	85.35	97.08	2.91	11.74

The analysis of the performance measures provided by RBFNN indicates that the expressions angry, happy, neutral were classified with CA greater than 96% from JAFFE dataset. Similarly the expressions fear, happy, sad and surprise were classified accurately with CA greater than 80% from KDEF dataset. Similarly, the model also provides the specificity (average) of 96.70% and 90.05% for JAFFE and KDEF datasets respectively which is an indication of lower false positives. The evaluation of type-I and type-II errors indicates the performance of the model in providing the least error values, FAR of 3.28% and FRR 2.38%, for JAFFE dataset. The two measures are 9.93% and 10.22% respectively for KDEF dataset.

TABLE 5. Performance metrics computed from testing dataset for MLPNN with $FV_{int\_norm\_2}$.

	JAFFE				KDEF			
Emotions	CA(%)	TNR(%)	FAR(%)	FRR(%)	CA(%)	TNR(%)	FAR(%)	FRR(%)
Angry	96.05	96.96	3.03	0.92	86.42	92.5	7.5	6.08
Disgust	90.78	95.38	4.61	4.61	85.71	90.79	9.2	5.09
Fear	88.15	95.31	4.68	7.17	91.07	97.91	2.08	6.85
Happy	96.05	96.87	3.12	0.83	88.92	95.41	4.5	6.58
Neutral	97.36	98.48	1.51	1.13	83.92	89.16	10.8	5.28
Sad	93.42	93.84	6.15	0.43	82.5	85	15	2.5
Surprise	98.68	98.48	1.51	0	89.64	95.83	4.16	6.2

TABLE 6. Performance metrics computed from testing dataset for RBFNN with $FV_{int\_norm\_2}$.

	JAFFE				KDEF			
Emotions	CA(%)	TNR(%)	FAR(%)	FRR(%)	CA(%)	TNR(%)	FAR(%)	FRR(%)
Angry	98.68	100	0	1.32	88.92	94.58	5.41	5.61
Disgust	93.42	96.92	3.07	3.51	88.21	92.91	7.08	4.71
Fear	92.1	96.87	3.12	4.78	89.64	95	5	5.36
Happy	97.4	98.43	1.56	1.04	92.14	97.5	2.5	5.36
Neutral	96.05	96.92	3.95	0	86.42	90.41	9.58	4
Sad	96.05	98.46	1.53	2.44	86.42	88.33	11.66	1.92
Surprise	97.36	96.96	2.64	0	90.35	95.83	4.16	5.49

2) CASE (II)

The experiment on subject dependent FER was continued with the next set of feature vectors. At this point, the texture features remain the same as considered in case (i) whereas the SPZM are captured by varying the ZM order from $n = 2$ to 4. The variation in n with the corresponding repetition factors resulted in $[|A_{20}|, |A_{22}|, |A_{31}|, |A_{33}|, |A_{40}|, |A_{42}|]$ which are utilized in equation (4) to provide SPZM of length 126. The order of ZM was varied to explore the competency of higher order ZM in presenting best shape features. Now these features are combined with LTexM features to form integrated feature vector $FV_{int\_norm\_2}$ with dimension 1×144.

The $FV_{int\_norm\_2}$ framed for the training datasets of JAFFE and KDEF are provided to both MLPNN and RBFNN classifiers. The classifiers were trained by varying the trainable parameters of the network to improve the performance. The MLPNN parameters optimized for $\eta = 0.5$, $m = 0.95$ and $\lambda = 0.6$ and RBFNN converged for $\sigma = 18$. The models were later evaluated for their performance by providing the samples of the testing data set. The measures computed on testing dataset with $FV_{int\_norm\_2}$ for both MLPNN and RBFNN are presented in Table 5 and Table 6.

The result and the analysis of the performance measures obtained from the second set of feature vectors $FV_{int\_norm\_2}$ implies improvement in the performance of the FER system. The average classification accuracies of 94.35% & 95.86% for JAFFE and 94.35% & 95.86% were achieved from MLPNN and RBFNN classifiers respectively. Another noteworthy improvement is in terms of type-I and type-II errors. The results clearly indicate that the average type-I error for all the expressions is reduced by 30% and type-II error by 36.84%. On the similar terms, CA of 86% and 88.87% was observed.

The results presented by MLPNN and RBFNN for both $FV_{int\_norm\_1}$ and $FV_{int\_norm\_2}$ were also examined to check for the discrimination ability. The assessment

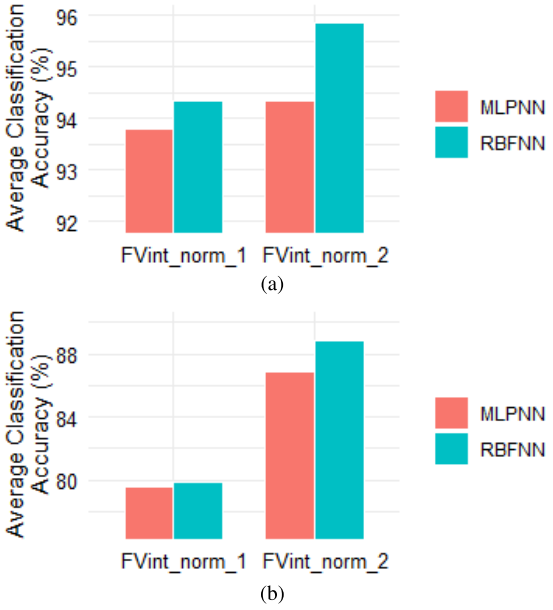

FIGURE 11. Average Classification Accuracy of the seven expressions (a) JAFFE and (b) KDEF databases.

TABLE 7. Performance comparison.

Methods	Average Classification/ Recognition accuracy (%)	
	JAFFE Dataset	KDEF Dataset
Histogram of oriented gradient(HOG) with SVM [50]	76.19	80.95
SURF with logistic regression[51]	-	74.05
VGG-16 with transfer learning, ResNet50 with transfer learning[52]	-	73.6, 76
DeepPCA with ELM[53]	-	83
Auxiliary Models [54]	95	-
LBP[55]	93.8	-
Adapted gradient LPQ[56]	92.59	-
Geometric features with modified HMM[57]	84.7	-
Histogram of oriented gradient(HOG) in wavelet domain[58]	71.43	-
Exemplar based[59]	92.53	-
Deep Convolutional Neural Network[60]	-	59.64
Kernel collaboration representation and deep subspace learning[61]	68.80	80.20
Modified LDP Features with the Optimization based Multi-SVNN Classifier[62]	96.00	-
Proposed method (RBFNN)	95.86	88.87

signified the performance of RBFNN that can be noted from the Fig. 11.

Later, the analysis is extended to compare the performance of the proposed method with state of the art facial expression recognition techniques and the same is presented in Table 7.

A comparative evaluation was carried out considering different image descriptors for the JAFFE and KDEF databases. The Table 7 provides the average classification accuracy which implies that the proposed method is comparable with the state of the art techniques.

TABLE 8. Performance metrics computed for subject independent FER with $FV_{int\_norm\_1}$.

	MLPNN				RBFNN			
Emotions	CA(%)	TNR(%)	FAR(%)	FRR(%)	CA(%)	TNR(%)	FAR(%)	FRR(%)
Angry	85.06	98.48	1.51	13.43	73.83	79.34	20.65	5.52
Disgust	76.16	95.38	4.61	19.23	80.64	94.5	5.49	13.87
Fear	71.49	95.31	4.68	23.83	74.29	82.06	17.93	7.78
Happy	70.56	95.31	4.68	24.76	81.77	92.53	7.06	11.17
Neutral	75.7	96.96	3.03	21.27	77.57	88.04	11.95	10.48
Sad	74.29	96.92	3.07	22.64	72.69	81.56	18.03	9.28
Surprise	82.24	93.47	6.52	11.24	82.24	89.69	10.32	7.44

TABLE 9. Performance metrics computed for subject independent FER with $FV_{int\_norm\_2}$.

	MLPNN				RBFNN			
Emotions	CA(%)	TNR(%)	FAR(%)	FRR(%)	CA(%)	TNR(%)	FAR(%)	FRR(%)
Angry	75.23	86.95	13.05	11.72	79.43	93.29	6.7	13.87
Disgust	69.15	78.02	21.98	8.87	75.23	86.81	13.18	11.59
Fear	79.43	90.76	9.23	11.34	78.03	88.04	11.95	10.02
Happy	74.76	84.69	15.31	9.93	73.83	81.42	18.57	7.6
Neutral	80.84	92.39	7.61	11.55	72.89	77.17	22.82	4.29
Sad	74.29	83.06	16.94	8.77	85.04	98.9	1.09	13.87
Surprise	75.23	84.23	15.77	9	81.77	92.93	2.17	16.06

B. SUBJECT INDEPENDENT FER

In the second experiment, the suitability of the extracted and integrated feature sets was checked for subject independent FER. To evaluate the performance, the combination of both JAFFE and KDEF databases was considered. The integrated features $FV_{int\_norm\_1}$ and $FV_{int\_norm\_2}$ were extracted from the facial images of both KDEF and JAFFE databases. The feature vectors of the KDEF dataset were labeled to form the training set while the JAFFE dataset forms the testing set. The performance metrics obtained with MLPNN and RBFNN are shown in Table 8 and Table 9.

The results were analyzed to find that the model with $FV_{int\_norm\_1}$ and MLPNN delivered a CA greater than 80 percentage for the angry and surprise expressions but for disgust, happy and surprise RBFNN delivers. Similarly, for the feature set $FV_{int\_norm\_2}$, an accuracy greater than 80 percentage was obtained for neutral expression with MLPNN and sad & surprise expressions with RBFNN. Also the average specificity is found to be greater than 85 percentage for both the integrated feature vectors for MLPNN and RBFNN that reduces the false positives. The metrics FAR and FRR presented in Tables 8 and 9 suggest that the proposed method has better generalization capability. The proposed method is able to identify the true facial expression independent to the subject. It is observed that the recognition of facial expression is improved by considering higher order ZM ($FV_{int\_norm\_2}$) and RBFNN performs better as compared to MLPNN in identifying the subject independent facial expression.

The descriptors were also significant in discriminating the emotions across the datasets (subject independent).

C. ABLATION STUDY

An additional experiment was conducted using ablation study to prove the effectiveness of integrating the Law's texture features with SPZM features in enhancing the performance of both subject dependent and independent FER.

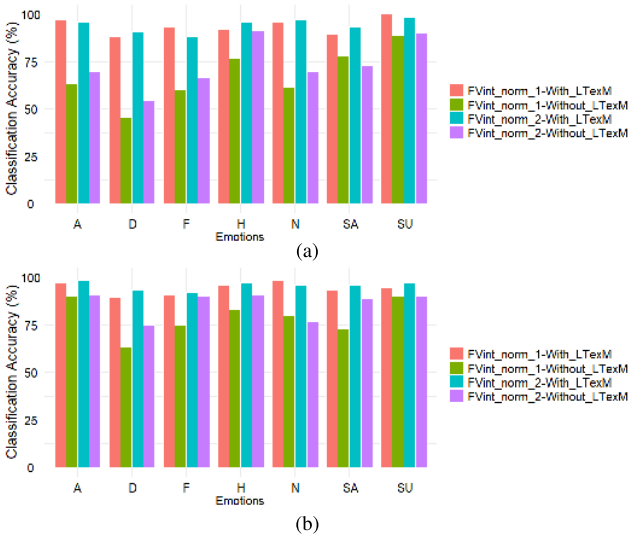

FIGURE 12. Plot of classification accuracies of all the expressions testing JAFFE dataset for subject dependent FER (a) MLPNN and (b) RBFNN.

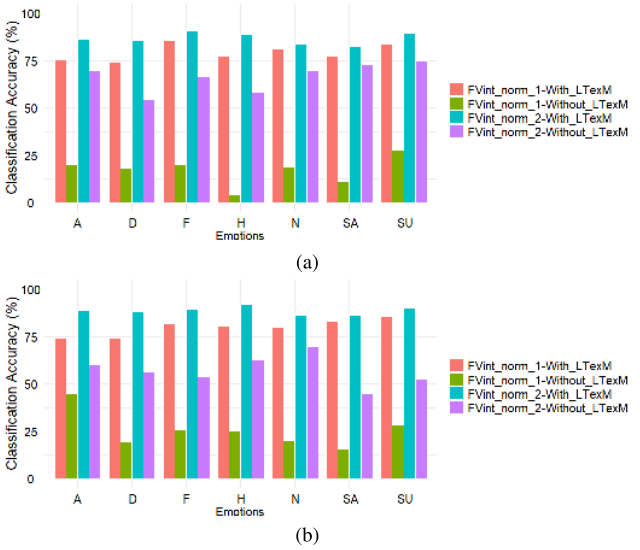

FIGURE 13. Plot of classification accuracies of all the expressions testing KDEF dataset for subject dependent FER (a) MLPNN and (b) RBFNN.

For ablation study, the Law's texture features were removed from the integrated feature set retaining only the SPZM features in both training set and testing set. $(FV_{int\_norm\_i}) = \{\text{SPZM}\}_{j=1(n=1,2,3),j=2(n=2,3,4)}$

Now the reduced training feature set with the class labels were provided to both MLPNN and RBFNN classifiers to create the model. For training the neural network the same criteria and tuning parameters were retained as mentioned in section III.D. The trained model is then tested with a reduced testing feature set. The performance is evaluated for recognition accuracy to investigate how the removal of texture features affects the performance of the subject dependent and independent FER. The results of the ablation study are displayed in Fig. 12, 13 and 14.

The classification accuracies shown in Fig. 12,13 and 14 obtained with the two neural network classifiers for both

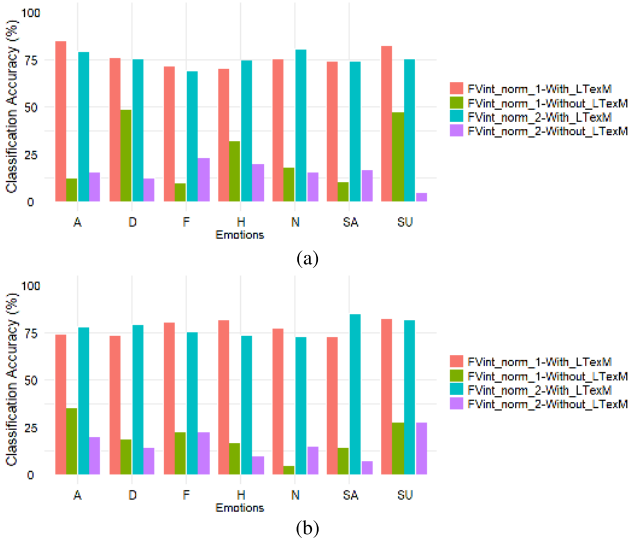

FIGURE 14. Plot of classification accuracies of all the expressions testing JAFFE dataset for subject independent FER (a) MLPNN and (b) RBFNN.

integrated feature vectors indicate that the combination of texture and shape descriptors are appropriate to represent the facial expressions for recognition followed by classification. From this study, It is noted that the integration of texture features contributes significantly in improving the recognition accuracy. It can also be noted that the RBFNN generalizes well by providing better performance in FER.

V. CONCLUSION

This paper presents a FER method using the combination of Law's texture features and ZM based shape descriptors. The proposed method uses feed forward neural networks MLPNN and RBFNN algorithms to learn the relationship between the features and the class labels of the emotions. The results proved that the proposed method has a good class discrimination ability and performed well with $FV_{int\_norm\_2}$ where the texture features were integrated with SPZM obtained from higher order ZM. The RBFNN classifier presented a noticeable result with an average recognition accuracy of 95.86% and 88.87% on the JAFFE and KDEF datasets respectively for subject dependent FER using RBFNN. The model classified the angry, happy and surprise expressions with good recognition rate whereas the performance for disgust, neutral and sad expressions is quite small in comparison. The method is also able to perform well even in the presence of facial orientations. Subsequently, the proposed method has better generalization capability in identifying the facial expressions across datasets for subject independent FER. The framed feature set proved to be accurate in capturing the facial expression patterns and thus returned less number of miss classifications by providing least possible type-I and type-II errors.

Further, the effectiveness of combined features can be extended for better understanding and recognition of the micro expressions. These expressions last only for a fraction of a second on the face and that actually represent the true feeling of an individual. The proposed method will be revised for both spatial and motion information of the image sequence to describe the key characteristics from the dynamic facial expressions for classification.

REFERENCES

[1] P. Ekman and W. V. Friesen, *Facial Action Coding System: Investigator's Guide*. Palo Alto, CA, USA: Consulting Psychologists Press, 1978.
[2] X. Fan and T. Tjahjadi, "A dynamic framework based on local zernike moment and motion history image for facial expression recognition," *Pattern Recognit.*, vol. 64, pp. 399–406, Apr. 2017.
[3] Y.-I. Tian, T. Kanade, and J. F. Cohn, "Recognizing action units for facial expression analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 2, pp. 97–115, Feb. 2001.
[4] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu, "Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron," in *Proc. 3rd IEEE Int. Conf. Autom. Face Gesture Recognit.*, Apr. 1998, pp. 454–459.
[5] A. L. A. Ramos, B. G. Dadiz, and A. B. G. Santos, "Classifying emotion based on facial expression analysis using Gabor filter: A basis for adaptive effective teaching strategy," in *Computational Science and Technology*. Singapore: Springer, 2020, pp. 469–479.
[6] E. Owusu, Y. Zhan, and Q. R. Mao, "A neural-AdaBoost based facial expression recognition system," *Expert Syst. Appl.*, vol. 41, no. 7, pp. 3383–3390, Jun. 2014.
[7] A. Hernandez-Matamoros, A. Bonarini, E. Escamilla-Hernandez, M. Nakano-Miyatake, and H. Perez-Meana, "Facial expression recognition with automatic segmentation of face regions using a fuzzy based classification approach," *Knowl.-Based Syst.*, vol. 110, pp. 1–14, Oct. 2016.
[8] G. P. Hegde, M. Seetha, and N. Hegde, "Kernel locality preserving symmetrical weighted Fisher discriminant analysis based subspace approach for expression recognition," *Eng. Sci. Technol., Int. J.*, vol. 19, no. 3, pp. 1321–1333, 2016.
[9] T. Ojala, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans.Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Aug. 2002.
[10] M. J. Cossetin, J. C. Nievola, and A. L. Koerich, "Facial expression recognition using a pairwise feature selection and classification approach," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 5149–5155.
[11] X. Wei, H. Wang, G. Guo, and H. Wan, "A general weighted multi-scale method for improving LBP for face recognition," in *Proc. Int. Conf. Ubiquitous Comput. Ambient Intell.*, Dec. 2014, pp. 532–539.
[12] M. Guo, X. Hou, Y. Ma, and X. Wu, "Facial expression recognition using ELBP based on covariance matrix transform in KLT," *Multimedia Tools Appl.*, vol. 76, no. 2, pp. 2995–3010, 2017.
[13] X. H. Guo, X. Zhang, C. Deng, and J. Wei, "Facial expression recognition based on independent component analysis," *J. Multimedia*, vol. 8, no. 4, pp. 402–409, 2013.
[14] S. Shah and D. Shah, "Partition-based face recognition using LDP and SVM," *J. Inf. Commun. Technol. Robot. Appl.*, vol. 8, no. 1, pp. 37–43, 2017.
[15] Y. Wu and W. Qiu, "Facial expression recognition based on improved local ternary pattern and stacked auto-encoder," *AIP Conf. Proc.*, vol. 1864, Jul. 2017, Art. no. 020131.
[16] B. R. Ilyas, B. Mohammed, M. Khaled, A. T. Ahmed, and A. Ihsen, "Facial expression recognition based on DWT feature for deep CNN," in *Proc. 6th Int. Conf. Control, Decis. Inf. Technol. (CoDIT)*, Apr. 2019, pp. 344–348.
[17] Z. Wang, Q. Ruan, and G. An, "Facial expression recognition using sparse local Fisher discriminant analysis," *Neurocomputing*, vol. 174, pp. 756–766, Jan. 2016.
[18] H. K. Meena, K. K. Sharma, and S. D. Joshi, "Effective curvelet-based facial expression recognition using graph signal processing," *Signal, Image Video Process.*, vol. 14, no. 2, pp. 241–247, Mar. 2020.
[19] M. Dahmane and J. Meunier, "Emotion recognition using dynamic grid-based HoG features," *Proc. Int. Conf. Autom. Face Gesture*, Mar. 2011, pp. 884–888.
[20] M. M. F. Donia, A. A. A. Youssif, and A. Hashad, "Spontaneous facial expression recognition based on histogram of oriented gradients descriptor," *Comput. Inf. Sci.*, vol. 7, no. 3, pp. 31–37, Jul. 2014.

[21] D. Ghimire and J. Lee, "Geometric feature-based facial expression recognition in image sequences using multi-class AdaBoost and support vector machines," *Sensors*, vol. 13, no. 6, pp. 7714–7734, Jun. 2013.
[22] A. Sánchez, J. V. Ruiz, A. B. Moreno, A. S. Montemayor, J. Hernández, and J. J. Pantrigo, "Differential optical flow applied to automatic facial expression recognition," *Neurocomputing*, vol. 74, no. 8, pp. 1272–1282, Mar. 2011.
[23] C. Singh, N. Mittal, and E. Walia, "Face recognition using zernike and complex zernike moment features," *Pattern Recognit. Image Anal.*, vol. 21, no. 1, pp. 71–81, Mar. 2011.
[24] M. Saaidia, N. Zermi, and M. Ramdani, "Facial expression recognition using neural network trained with zernike moments," in *Proc. 4th Int. Conf. Artif. Intell. with Appl. Eng. Technol.*, Dec. 2014, pp. 187–192.
[25] M. Ahmady, R. Ghasemi, and H. Rashidy Kanan, "Local weighted pseudo zernike moments and fuzzy classification for facial expression recognition," in *Proc. 13th Iranian Conf. Fuzzy Syst. (IFSC)*, Aug. 2013, pp. 1–4.
[26] R. Jafri and H. R. Arabnia, "A survey of face recognition techniques," *J. Inf. Process. Syst.*, vol. 5, no. 2, pp. 41–68, 2009.
[27] V. G. Mahesh, A. N. Raj, and Z. Fan, "Invariant moments based convolutional neural networks for image analysis," *Int. J. Comput. Intell. Syst.*, vol. 10, no. 1, pp. 936–950, 2017.
[28] N. Zeng, H. Zhang, B. Song, W. Liu, Y. Li, and A. M. Dobaie, "Facial expression recognition via learning deep sparse autoencoders," *Neurocomputing*, vol. 273, pp. 643–649, Jan. 2018.
[29] K. Zhang, Y. Huang, Y. Du, and L. Wang, "Facial expression recognition based on deep evolutional spatial-temporal networks," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4193–4203, Sep. 2017.
[30] Y. Liu, X. Yuan, X. Gong, Z. Xie, F. Fang, and Z. Luo, "Conditional convolution neural network enhanced random forest for facial expression recognition," *Pattern Recognit.*, vol. 84, pp. 251–261, Dec. 2018.
[31] S. Xie and H. Hu, "Facial expression recognition using hierarchical features with deep comprehensive multipatches aggregation convolutional neural networks," *IEEE Trans. Multimedia*, vol. 21, no. 1, pp. 211–220, Jan. 2019.
[32] K. Li, Y. Jin, M. W. Akram, R. Han, and J. Chen, "Facial expression recognition with convolutional neural networks via a new face cropping and rotation strategy," *Vis. Comput.*, vol. 36, no. 2, pp. 391–404, Feb. 2020.
[33] H. Zhang, B. Huang, and G. Tian, "Facial expression recognition based on deep convolution long short-term memory networks of double-channel weighted mixture," *Pattern Recognit. Lett.*, vol. 131, pp. 128–134, Mar. 2020.
[34] X. Wang, X. Wang, and Y. Ni, "Unsupervised domain adaptation for facial expression recognition using generative adversarial networks," *Comput. Intell. Neurosci.*, vol. 2018, Jul. 2018, Art. no. 7208794.
[35] H. Yang, Z. Zhang, and L. Yin, "Identity-adaptive facial expression recognition through expression regeneration using conditional generative adversarial networks," in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit.*, May 2018, pp. 294–301.
[36] *The Japanese Female Facial Expression (JAFFE) Dataset*. Accessed: Feb. 23, 2016. [Online]. Available: http://www.kasrl.org/jaffe.html
[37] D. Lundqvist, A. Flykt, and A. Öhman, "The Karolinska directed emotional faces (KDEF)," Dept. Clin. Neurosci., Psychol. Sect., CD ROM From, Karolinska Inst., Solna, Sweden, 1998, vol. 91, no. 630, p. 2.
[38] A. Khotanzad and Y. H. Hong, "Invariant image recognition by zernike moments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 5, pp. 489–497, May 1990.
[39] A. Tahmasbi, F. Saki, and S. B. Shokouhi, "Classification of benign and malignant masses based on zernike moments," *Comput. Biol. Med.*, vol. 41, no. 8, pp. 726–735, Aug. 2011.
[40] V. G. Mahesh and A. N. J. Raj, "Invariant face recognition using Zernike moments combined with feed forward neural network," *Int. J. Biometrics*, vol. 7, no. 3, pp. 286–307, 2015.
[41] V. Mahesh, A. Raj, and T. Celik, "Silkworm cocoon classification using fusion of zernike moments-based shape descriptors and physical parameters for quality egg production," *Int. J. Intell. Syst. Technol. Appl.*, vol. 16, no. 3, pp. 246–268, 2017.
[42] B.-D. Liu, J. Meng, W.-Y. Xie, S. Shao, Y. Li, and Y. Wang, "Weighted spatial pyramid matching collaborative representation for remote-sensing-image scene classification," *Remote Sens.*, vol. 11, no. 5, p. 518, Mar. 2019.
[43] S. Lazebnik, C. Schmid, and J. Ponce, "Spatial pyramid matching," *Object Categorization, Comput. Human Vis. Perspect.*, vol. 3, no. 4, p. 5, 2009.
[44] K. I. Laws, "Texture energy measures," in *Proc. Image Understanding Workshop*, Nov. 1979, pp. 47–51.
[45] O. Faust, U. R. Acharya, K. M. Meiburger, F. Molinari, J. E. W. Koh, C. H. Yeong, P. Kongmebhol, and K. H. Ng, "Comparative assessment of texture features for the identification of cancer in ultrasound images: A review," *Biocybern. Biomed. Eng.*, vol. 38, no. 2, pp. 275–296, 2018.
[46] A. S. Setiawan, Elysia, J. Wesley, and Y. Purnama, "Mammogram classification using Law's texture energy measure and neural networks," *Procedia Comput. Sci.*, vol. 59, pp. 92–97, Jan. 2015.
[47] S. Dash and U. R. Jena, "Texture classification using steerable pyramid based Laws' masks," *J. Electr. Syst. Inf. Technol.*, vol. 4, no. 1, pp. 185–197, May 2017.
[48] M. T. Hagan and M. B. Menhaj, "Training feedforward networks with the marquardt algorithm," *IEEE Trans. Neural Netw.*, vol. 5, no. 6, pp. 989–993, 1994.
[49] J. Han, J. Pei, and M. Kamber, *Data Mining: Concepts and Techniques*. Amsterdam, The Netherlands: Elsevier, 2011.
[50] S. K. Eng, H. Ali, A. Y. Cheah, and Y. F. Chong, "Facial expression recognition in JAFFE and KDEF Datasets using histogram of oriented gradients and support vector machine," in *Proc. IOP Conf. Ser., Mater. Sci. Eng.*, Nov. 2019, vol. 705, no. 1, pp. 1–7.
[51] Q. Rao, X. Qu, Q. Mao, and Y. Zhan, "Multi-pose facial expression recognition based on SURF boosting," in *Proc. Int. Conf. Affect. Comput. Intell. Interact. (ACII)*, Sep. 2015, pp. 630–635.
[52] Z. Elgarrai, O. E. Meslouhi, M. Kardouchi, and H. Allali, "Robust facial expression recognition system based on hidden Markov models," *Int. J. Multimedia Inf. Retr.*, vol. 5, no. 4, pp. 229–236, Nov. 2016.
[53] K. Rujirakul and C. So-In, "Histogram equalized deep PCA with ELM classification for expressive face recognition," in *Proc. Int. Workshop Adv. Image Technol. (IWAIT)*, Jan. 2018, pp. 1–4.
[54] Y. Wang, Y. Li, Y. Song, and X. Rong, "Facial expression recognition based on auxiliary models," *Algorithms*, vol. 12, no. 11, p. 227, Oct. 2019.
[55] M. Pietikäinen and A. Hadid, "Texture features in facial image analysis," in *Proc. Int. Workshop Biometric Person Authentication*. Berlin, Germany: Springer, Oct. 2005, pp. 1–8.
[56] S. Kherchaoui and A. Houacine, "Facial expression identification using gradient local phase," *Multimedia Tools Appl.*, vol. 78, no. 12, pp. 16843–16859, Jun. 2019.
[57] M. Rahul, "Facial expression recognition using geometric features and modified hidden Markov model," *Int. J. Grid Utility Comput.*, vol. 10, no. 5, pp. 488–496, 2019.
[58] S. Nigam, R. Singh, and A. K. Misra, "Efficient facial expression recognition using histogram of oriented gradients in wavelet domain," *Multimedia Tools Appl.*, vol. 77, no. 21, pp. 28725–28747, Nov. 2018.
[59] N. Farajzadeh and M. Hashemzadeh, "Exemplar-based facial expression recognition," *Inf. Sci.*, vols. 460–461, pp. 318–330, Sep. 2018.
[60] Y. I. Colon, M. Q. Hill, C. J. Parde, and C. D. Castillo, "Facial expression information in deep convolutional neural networks trained for face identification," *J. Vis.*, vol. 19, no. 10, p. 93, 2019.
[61] Z. Sun, Z.-P. Hu, R. Chiong, M. Wang, and W. He, "Combining the kernel collaboration representation and deep subspace learning for facial expression recognition," *J. Circuits, Syst. Comput.*, vol. 27, no. 8, Jul. 2018, Art. no. 1850121.
[62] I. Michael Revina and W. R. Sam Emmanuel, "Face expression recognition with the optimization based multi-SVNN classifier and the modified LDP features," *J. Vis. Commun. Image Represent.*, vol. 62, pp. 43–55, Jul. 2019.

VIJAYALAKSHMI G. V. MAHESH received the B.E. degree in electronics and communication engineering from Bangalore University, India, in 1999, the M.Tech. in digital communication and networking from Visvesvaraya Technological University, in 2005, and the Ph.D. degree from the Vellore Institute of Technology, Vellore, India. She is currently working as an Associate Professor with the BMS Institute of Technology and Management, Bengaluru, India. Her research interests include pattern recognition, machine learning, image processing, and deep learning.

CHENGJI CHEN was born in Guangdong, China. He received the Bachelor of Engineering degree, in 2019. He is currently pursuing the master's degree in engineering with Shantou University, Shantou, China. His research interests include image processing and facial expression recognition.

VIJAYARAJAN RAJANGAM (Member, IEEE) received the Bachelor of Engineering degree in electronics and communication from the University of Madras, in 1998, and the master's degree in applied electronics from Madurai Kamarajar University, in 1999. He did his Ph.D. research in image fusion with Anna University, in 2015. He is currently working as a Faculty Member with the Centre for Healthcare Advancement, Innovation and Research, School of Electronics Engineering, VIT University, Chennai. He has been in academics for over 20 years and has published his research in various reputed journals and conferences. His research interests include image fusion, bio-cryptography, emotion classification using signal and image processing, and deep learning for image analysis.

ALEX NOEL JOSEPH RAJ (Member, IEEE) received the B.E. degree in electrical engineering from Madras University, India, in 2001, the M.E. degree in applied electronics from Anna University, in 2005, and the Ph.D. degree in engineering from the University of Warwick, in 2009. From October 2009 to September 2011, he was a Design Engineer with Valeport Ltd., Totnes, U.K. From March 2013 to December 2016, he was a Professor with the Department of Embedded Technology, School of Electronics Engineering, Vellore Institute of Technology, Vellore, India. Since January 2017, he has been with the Department of Electronic Engineering, College of Engineering, Shantou University, China. His research interests include deep learning, signal and image processing, and FPGA implementations.

PALANI THANARAJ KRISHNAN (Member, IEEE) received the Ph.D. degree from the Faculty of Information and Communication Engineering, Anna University, Chennai, in 2018. He is currently working as an Assistant Professor with the Department of Electronics and Instrumentation Engineering, St. Joseph's College of Engineering, Chennai. He has developed deep learning algorithms for performing image classification of medical images for disease diagnosis. He has published his works in many reputed and refereed journals indexed in Web of Science and Scopus. His research interests include image processing, advanced signal processing, image segmentation, machine learning, and deep learning.