



# Emotion estimation from nose feature using pyramid structure

S.G. Shaila<sup>1</sup> · A Vadivel<sup>2</sup> · Suma Avani<sup>1</sup>

Received: 16 December 2021 / Revised: 22 March 2022 / Accepted: 3 February 2023 /

Published online: 15 April 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

## Abstract

The facial expression and human emotions are considered as important components for building various real-time applications such as human expression and emotion recognition systems. Various parts of the human face contribute to recognizing expression. The contribution of action units on the nose is considered important. In this paper, input images are converted into HSV color space for better representation. The nose area is localized and the boundary is drawn by segmentation process using Fuzzy C-means Clustering (FCM). The segmented nose on the human face is modelled using a pyramid/tetrahedron structure and it is superimposed on the reference face. The feature points are identified on the pyramid model, where the Action Units (AUs) falling on the tetrahedron are identified. These points are validated with the theoretical properties of the tetrahedron so that the constructed feature vector is robust. The degree of deformation at various points is constructed as the feature vector. The feature vector is extracted for all the database images, say JAFE and CK++ datasets, and the feature database is created and stored separately. The feature data sets are used for training and thus, they are n-fold cross-validated to avoid over and under fitting. Given an input image for estimating the expression and emotion, the feature vector of the input image is compared with the feature vector of deformed images stored in the database. We have used Support Vector Machine (SVM), and Multilayer Perceptron (MLP) and Random Forest classifier to classify the expression and derive emotion. The JAFE and CK++ datasets are used for experimental analysis. It is found that the Nose feature using pyramid/tetrahedron structure is giving

---

✉ A Vadivel  
vayyasam@gitam.edu

S.G. Shaila  
shaila-cse@dsu.edu.in

Suma Avani  
sumaavani2@gmail.com

<sup>1</sup> Department of CSE, Dayananda Sagar University, Bangalore, India

<sup>2</sup> Department of Computer Science and Engineering, GITAM School of Technology, GITAM University, Bengaluru Campus, Bengaluru, India

good results. Most of the time the classification accuracy is more than 95%. The experimental results are compared with some of the well-known approaches and the proposed tetrahedron model has performed well with classification accuracy more than 95%.

**Keywords** Facial expression · Human emotions · Nose structure · Geometric feature · Tetrahedron and pyramid · Triangular shape · Deformation · Classification

## 1 Introduction

In general, facial expression plays a major role in signifying the emotional and mental state of a human being. Psychologists say that 7% of the total information is conveyed through language - 38% through language auxiliaries such as speech rhythm, tone, etc., and 55% through facial expression. Hence, Facial Expression Recognition (FER) is considered an important research topic for real-time life applications such as computer vision, robotics, medical treatment, driver fatigue surveillance, and much other Human-Computer Interaction (HCI) applications. Ekman and Friesen [16] have proposed 6 basic emotions from the human facial images such as *anger*, *happiness*, *surprise*, *sadness*, *disgust*, and *fear*, that are expressed frequently through facial expressions. These can occur by the events which are expressed during communication. The traditional approaches may not handle the requirement of real-time applications effectively. Also, the effect of change in illumination conditions, noise, occlusion, and variations in head posture are some of the important factors that affect the expression recognition systems resulting in uncertainty and ambiguity. Many techniques use handcrafted features for facial expression recognition and thus, require increased efforts in terms of programming and computation cost.

It is well-known that expression and human emotions are natural and intuitive processes. According to Psychologists, various parts of the human face contribute to different categories of emotions. For instance, the upper part of the face such as the eyes and eyebrows contribute to *anger* and *fear*. On the other hand, *disgust* and *happiness* are mainly expressed through the mouth [10], and *surprise* is influenced by the upper and lower parts of the face. In complex applications such as Human-Computer Interaction (HCI), assistive robotics, Gaming, etc., it is important to measure the expression with accurate emotional information having minimum error [9].

It is observed that automatic FER is a complex task due to variations in the features of faces in terms of a person's identity, illumination conditions, head pose, etc. Presently available human-computer interfaces are unable to provide natural interaction to the users as they are not taking complete advantage of the communicative media. The HCI can be improved significantly, provided computer systems recognize the user's emotions through facial expressions, and body gestures and react to the users' wishes. Thus, the researchers have shown interest in automatic FER systems as the classical systems have failed miserably.

It is known that emotion recognition deals with two primary steps such as feature extraction and classification. Feature extraction refers to the selection of a set of features or attributes that influences emotional expressions. The emotional expression in turn is influenced by the motions and positions of the facial features and their muscles beneath the skin. These feature movements convey the emotional state to an observer. The classification mechanism handles the classification scheme for identifying the class of emotion. As a result, FER systems can be evaluated based on recognition rate and is purely dependent on the feature extraction

strategies. At times, a good classifier may not provide high classification accuracy if the extracted features are poor. On the other hand, for a large set of features that describes an emotion, the recognition system fails due to a poor classifier. Thus, many techniques are used for feature selection such as Principal Component Analysis (PCA), Independent Component Analysis (ICA), Rough Sets, Gabor Filter, Fourier Descriptors, etc. Similarly, various machine learning techniques such as Neural Networks, Fuzzy approaches, Support Vector Machine (SVM), Hidden Markov Model (HMM), etc. are used for facial expression classification. In addition, various segmentation algorithms have also been proposed recently and one of them is [37, 38] proposed automating the segmentation process of brain images.

Thus, the paper deals with the above-discussed issues and considers the nose feature of the human face, and tries to understand its proportion of contribution in terms of facial expressions. To the best of our knowledge, no literature is available, that uses the pyramid structure to extract the feature from the nose. The sample images are shown with handcrafted drawings for better clarity. Otherwise, the method uses an automated way to extract the features of the nose. We extract the feature (micro-level) of the nose through segmentation, localizing the nose, etc. The advantage of the proposed feature as compared with the deep learning feature is that we use a geometrical model based on a pyramid and tetrahedron. It is robust such that it can be used on any type of nose. However, while using deep learning, we have to propose more numbers of filters to model the structure of the nose, which is complex and time-consuming.

The proposed approach builds a geometrical model to map the human nose and superimpose on it a pyramid structure. The points of interest and their deformation are constructed from the input image, which is a Region of Interest (RoI), and used as a feature vector for further analysis and interpretation. The rest of the article is organized as follows:

In Section 2, the recent and relevant literature is reviewed.

In Section 3, the pyramid structure of the nose is proposed and a feature extraction scheme is presented.

In Section 4, the experimental results are presented with discussion and analysis.

We have concluded the paper in the last section.

## 2 Literature survey

The application of facial expression recognition from the human face has many folds such as computerized psychological counseling and therapy, identifying criminal and antisocial motives, etc. This section represents some of the recent state-of-the-art literature. In early 1975, the authors [15] have measured facial expressions with the help of the movement of the cheek, chin regions, and wrinkles of the human face. Various authors have proposed approaches such as Fuzzy Integral [26], Fourier descriptor [42], Template Matching [8], and Neural Network models [18] to measure the human expressions. Cohen et al [12] recognize emotions' temporal variations in facial expressions from the live video display. The frames of the videos are continuously monitored and the features were extracted frame-wise. However, it is noticed that measuring the facial expression from the video is hard as its resolution and clarity have been low. Gao et al. [20] have presented the line-based caricatures method and used the single facial image to recognize facial expression. The performance of this approach was found to be low as it has modelled the imitation of a person from the picture, and the time complexity of the

algorithm was poor. Li and Ji [28] have proposed an approach based on a probabilistic framework to dynamically model and recognize the user's expressions. It has assisted in the form of a response. However, one of the drawbacks of this approach is that segmenting the Region of Interest (RoI) for localizing the area was difficult. This is due to the fact that there is a significant variation in the attribute values of different regions of the face. The authors [54] have stated that facial expressions have a direct correlation to many basic movements of human eyes, eyebrows, and mouth. A fuzzy kernel clustering has been used to identify the correlation between basic movements of the facial action units. It is reported that it was hard to measure the exact facial expression from blurred input faces since this method was unable to handle the blurred images. Chakraborty et al. [11] have studied the psychological states of the subjects which affect the emotional features. For instance, the wider variations of facial expressions of the human face have influenced the individual features of emotion. There are differences in facial features of the different subjects for the same emotion. There is a small random variation in the facial feature around a specific feature point. This has been found in multiple human face instances of similar emotions by conducting repeated experiments. However, it is noticed that various subjects of the human face have many complexities in terms of features of respective facial components. Zhang and Liu, [51] have proposed fuzzy logic approach for recognizing the emotional expression from the human face. The emotion has been excited by audiovisual stimulus and the variation in the facial regions are recorded as video movie clips. It is observed that the automated emotion recognition system has failed to map human emotions with human perception in many of the video clips.

In recent times, automated segmentation techniques are used for segmenting objects in images. The authors in SivaSai et al. [36] have proposed an automated fuzzy-based segmentation technique and recurrent neural network. The image is preprocessed using the Adaptive Bilateral Filter (ABF) for removing noises. The binary thresholding and Fuzzy Recurrent Neural Network (FR-Net) have been used for segmenting Region of Interest (ROI). The authors in [39] have presented different optimization techniques that are used for automated segmentation procedures. The performance of algorithms has been evaluated quantitatively using the genetic algorithm-based approach. A two-level classification approach has been presented by Drume and Jalal [14]. Both Principal Component Analysis (PCA), and Support Vector Machine (SVM) have been used for classification. However, the accuracy has not been encouraging since the improvement has been marginal. Benitez-Quiroz et al. [4] have proposed an approach that annotates both the Action Units (AUs) and corresponding intensities to label the motion categories of human facial expression. Various statistical parameters of facial points such as distances and angles between facial landmarks, etc. have been defined as geometric features. The change in the shading feature on the skin regions has been extracted using Gabor filters to represent both the shape and geometric features of every AUs. The activation level of every AU is calculated using Kernel Subclass Discriminant Analysis (KSDA) [48] to determine emotion categories. A method has been proposed using a feature selection algorithm to classify emotions based on appearance [24]. However, this approach failed to handle noise and generalized the appearance model. Demir [13], has used Local Curvelet Transform (LCT) as a feature descriptor to extract the geometric features based on the wrapping mechanism. The extracted geometric features are mean, entropy, and standard deviation. In addition, geometrical features such as energy, and kurtosis are extracted using a three-stage steerable pyramid representation Mahersia and Hamrouni [31]. The author Biswas [7] has extracted features using Discrete Contourlet Transform (DCT) and it has been performed in two key stages such as Laplacian Pyramid (LP), and Directional Filter Bank

(DFB). In the LP stage, the image is portioned into the low pass, bandpass and confines the discontinuities position. In the DFB stage, the bandpass portion is processed to form a linear composition by associating the discontinuities' position. Siddiqi et al. [35] have used Stepwise Linear Discriminant Analysis (SWLDA) to extract the localized features with backward and forward regression models. However, this method has faced challenges to recognize the emotion while there is variation in terms of age, culture, and gender. It is well-known that variations in the size of the image, the facial orientation, glasses or masks on the faces, and lightning conditions are the factors that increase the complexity in FER. Vivek and Reddy [44] have combined three well-known machine learning algorithms such as Cat Swarm Optimization (CSO), Genetic Algorithm (GA), and Particle Swarm Optimization (PSO) for determining the human facial expression. However, the classification is poor while there is a fusion of emotions. Similarly, Alhoussein [1] has used a set of well-known algorithms such as Radon Transform (RT), and Higher-Order Spectral (HOS) and found that the misclassification rate was on the higher side. A large number of methods such as Biorthogonal Wavelet Entropy, Fuzzy SVM, and Stratified Cross-Validation have been combined to recognize human emotion in Zhang et al. [53]. However, the performance has failed due to outliers and noises.

Liu et al. [30] have constructed the features which are supposed to be informative and non-redundant to facilitate learning and generalization. The approach has used PCA and LBP for extracting features from active facial patches. However, the classification accuracy has been reduced due to the presence of variation in illumination. Kumar et al. [27] have proposed a Weighted Projection-based LBP (WPLBP) approach to extract features based on the instructive regions. The LBP are assigned weights based on the significance of the instructive regions. The recognition rate of this approach has been reduced due to occlusions and head pose variations. Salmam et al. [33] have proposed a Supervised Descent Method (SDM) for extracting features in three stages. In the first and second stages, the facial main positions are extracted for selecting the related positions. At the final stage, it estimates the distance between the various components of the face using Classification and Regression Tree (CART). The metric likely Decision tree and Gini impurity are estimated to classify the expression. However, this method has failed to perform well if the input image is a lower resolution one.

Li et al. [29] have created the RAF-DB database having 29,672 image instances that are based on the crowdsourcing annotation. These images are labeled independently with forty annotations.

A deep locality preserving learning algorithm has been proposed for recognizing facial expressions and achieved only 44.55% as accuracy. Ghimire et al. [22] have used fifty-two facial key points and these points are extended as lines and triangles from which the geometrical features are extracted. The triangular geometric features have achieved good accuracy and however, it has failed for low-resolution images. Zavarez et al. [49] have employed Gabor motion energy filters to identify the dynamic facial expressions of individuals. These filters along with Genetic Algorithms (GA) and SVM are used as a feature for experiments to classify the facial expressions from video sequences. One of the issues with the approach is the time complexity. Zhong et al. [55] have proposed a FER framework based on subspace learning on local structure. Bilkhu et al. [6] have proposed a FER system to recognize six emotional features. A cascaded Regression Tree is applied for extracting the features. The system has used Logistic Regression, Support Vector Machine, and Neural Network for classifying the six emotions. However, one of the issues of this approach is that it has failed in distinguishing certain expressions such as disgust with anger and sadness with fear. This is due to the fact that the high similarities between these classes of facial expressions

which leads to higher misclassification rate. Florea et al. [19] have proposed an approach that combines semi-supervised learning and inductive transfer learning into an Annealed Label Transfer (ALT) framework to tackle the label scarcity issue in FER. ALT transfers a learner's knowledge on a labeled wild FER dataset to an unlabeled face dataset to generate pseudo labels. These labels' confidence is increased to improve the primary learner's performance in recognition. Zeng et al. [50] have proposed Inconsistent Pseudo Annotations to Latent Truth (IPA2LT) to address the label noise and alleviate prediction bias to a specific wild FER dataset. IPA2LT trains a Latent Truth Network (LTNet) to extract the true latent label for a sample using the inconsistency between the labels generated with a prediction model and manual labels. Ghazouani [21], has proposed the GP – FER framework, genetic programming is used for selecting features, and fusion technique is applied for facial expression recognition. Tree-based genetic program with three functional layers (feature selection, feature fusion, and classification) are formed. The binary classifier is used and it performs discriminative feature selection and fusion for each pair of expression classes. The final emotion is captured by performing a unique tournament elimination between all the classes using the binary programs. Yao, et al. [47] have used active learning and Support Vector Machine (SVM) to classify facial action units (AU) for facial recognizing facial expression. Active learning was used to detect the targeted facial expression AUs, while SVM was utilized to classify different AUs and ultimately map them to their corresponding facial expressions. Shahid et al. [34] have proposed a two-stage approach to distinguish seven expressions on the basis of eleven different facial areas. The combination of contour and region harmonics is used to develop the interrelationship of sub-local areas in the human face. The authors have applied a multi-class support vector machine (SVM) with subject-dependent k-fold cross-validation to classify human emotions into expressions. The authors Swaminathan, and Vadivel, [41] have proposed 37 emotions as combined emotions in which 16 emotions are newly derived and validated by using the Facial Action Coding System (FACS) and statistical analysis of combined emotions. Avani et al. [2] have proposed the parabola theory to map with mouth features. The points on the lips are considered feature points to construct feature vectors. The Latus Rectum, Focal Point, Directrix, Vertex, etc. are considered to identify the feature points of the lower lips and upper lips to understand the facial expressions. Avani et al. [3] have proposed an Interval graph of facial regions for FER. Here the approach assumed that common intersecting salient points of facial regions for estimating the emotions. The facial region is decomposed into four sub-regions and the Interval graph is extracted for each region. The common salient points and degree and direction of deformation are measured for vertical, horizontal, and diagonal directions. Wang et al. [45] have addressed label uncertainty by proposing a Self-Cure Network (SCN) to re-label the mislabeled samples. A self-attention mechanism estimates a weight for each sample in a batch based on the network's prediction and identifies label uncertainty using a margin-based loss function. One of the drawbacks of this approach is that it has failed in recognizing the expressions while the face in the image is non-frontal, varying lighting conditions and occlusions.

The same authors further designed a Region Attention Network (RAN) to address pose and occlusion in wild FER datasets bypassing regions around facial landmarks to a CNN [46]. The final feature vector is obtained by combining weighted feature vectors of cropped regions using a self-attention module. Though the model learned to discriminate occluded and non-occluded faces, it has failed to distinguish disgust with anger and sadness with fear. Minaee et al. [32] have proposed a deep learning method based on the attentional convolutional network that focuses on critical areas of the face on various datasets, including FER-2013,

CK+, and JAFFE. It encodes shape, appearance, and extensive dynamic information. However, one of the issues of this approach is that it requires more time for training the data. A significant challenge that deep FER systems face is the lack of sufficient quality training data in large volumes. This issue has been handled by proposing feature extraction techniques by Fan et al. [17]. This approach hybridizes several machine learning methods, such as support vector regression (SVR), grey catastrophe (GC), and random forest (RF). However, it is found that this method requires voluminous training data to converge the training algorithm. In recent times, hypotheses are proposed and validated using statistical tests on time series data. Though these methods are effective for time series data, they are found to be unsuitable for the features of the proposed work.

Based on the above review, it is imperative that a method is required to handle various issues such as convergence of training and learning algorithm, variation in illumination, pose and orientation, using the importance of nose component to improve the performance. As a result, we propose a novel scheme to extract the feature from the nose-based properties of pyramid and tetrahedron. Since this feature is on geometrical properties, it handles the variation in illumination, pose, and orientation. The performance is measured in terms of classification accuracy and found that the misclassification rate is very low.

### 3 Proposed approach

#### 3.1 Segmentation of the nose region

The input image is converted as an HSV image for effectively processing the input. The HSV color space is more robust towards change in external lighting. In cases, there is a minor change in external lighting (such as pale shadows, etc.). Hue values vary relatively less as compared to RGB values. The colors used in HSV can be related to human perception [43], which is not always the case with RGB or CMYK. The image consists of three channels such as Hue, Saturation, and Value. This color model does not use primary colors directly. It uses color in the way humans perceive them. As per Stockman and Shapiro [40], HSV color space can be represented pictorially as hexacone as shown in Fig. 1. The Intensity (I) or Intensity Value (V) is the vertical axis, H is the color represented in angular values, and S denotes the saturation for measuring the depth of the color.

Below, we present the proposed approach that considers and analyses human expressions such as happy, sad, fear, anger, disgust, surprise, etc. We consider the human nose to analyze its contribution in the facial expression and map the tetrahedron/pyramid geometric structure to extract the important feature vectors. The overall flow diagram of the proposed approach is depicted in Fig. 2 below.

It is observed from Fig. 2 that given an input image, the nose component of the face is localized and segmented using the boundary information in the preprocessing stage. We have used the Fuzzy C-Means clustering for segmentation and corresponding mathematics are given in Eqs. 1–3. A sample output of the clustering is given below in Fig. 3.

The proposed geometrical structure based on the properties of the pyramid is superimposed over the nose component of the facial image after localizing the nose and the same is depicted in Fig. 3. In Fig. 3(a), the sample image from the JAFFE dataset is depicted and Fig. 3 (b) and (c) are segmented images and pyramid structure superimposed images. Similarly, Fig. (d), (e), and (f) are the original, segmented, and pyramid structure superimposed images of another sample

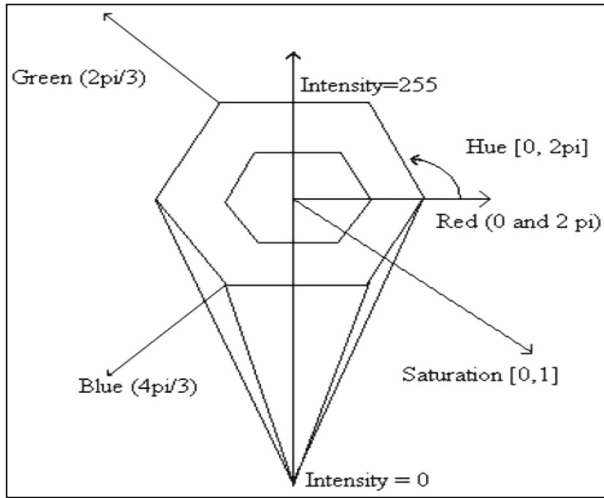


Fig. 1 The HSV color space – hexacone

image from the JAFE dataset. Various components such as eyes, mouth, etc. of the face are used as a reference for exactly superimposing the pyramid structure and an example is given in Fig. 4. The explanation of variables of the pyramid structure is explained later in the paper.

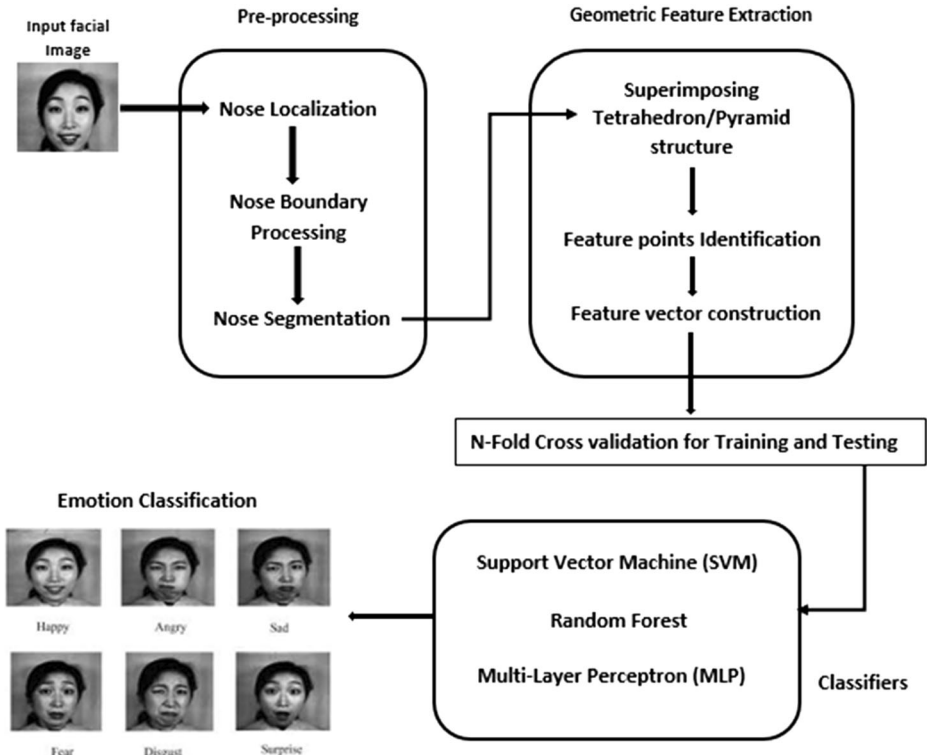
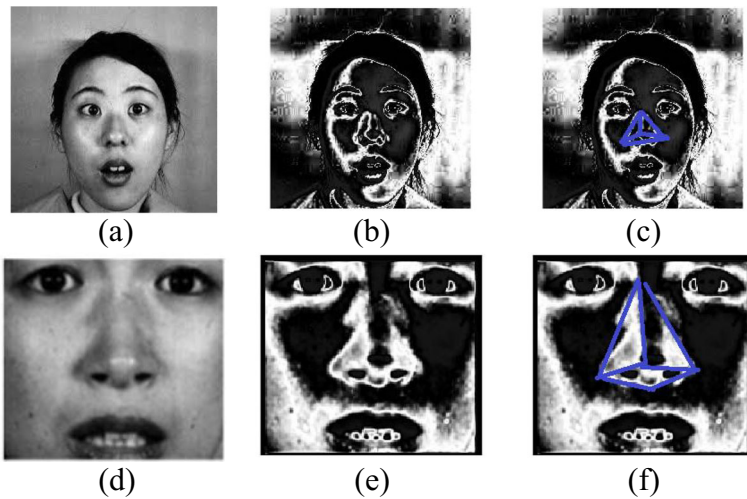


Fig. 2 The framework of the proposed approach



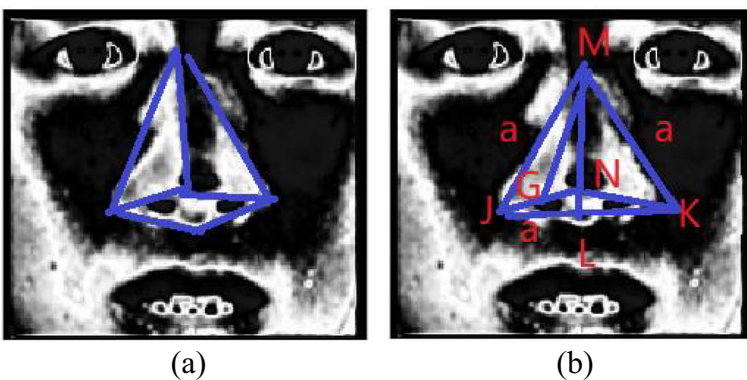


**Fig. 3** Sample output of Fuzzy C-Mean Clustering (a) Original Image (b) Segmented Image of Sample 1 from JAFE data set (c) Nose Localization of Sample 1 (d) Another Sample Image (e) Segmented Image of Sample 2 from JAFE data set (f) Nose Localization of Sample 2

The training and testing phase of the proposed work applies n-fold cross-validation and is depicted in Fig. 5. During the training phase, the feature is extracted for all the training images and stored in a repository. Similarly, the feature is extracted for a given input image and compared with all the features in the repository. We have used Manhattan distance for calculating the similarity value between the feature of query facial image and features stored in the repository. Since we have used n-fold cross-validation for all the features stored in the repository, all the features have contributed both in testing and training for improving the classification accuracy.

Below we present the mathematical base of the proposed approach.

The Fuzzy Clustering Algorithm (FCM) [5] is used for segmenting the nose region of the face. It is a clustering scheme that handles a piece of data that divides data into two or more clusters.



**Fig. 4** Nose component area is super imposed with pyramid structure. (a) Sample nose localized image (b) Feature Points (Variables)

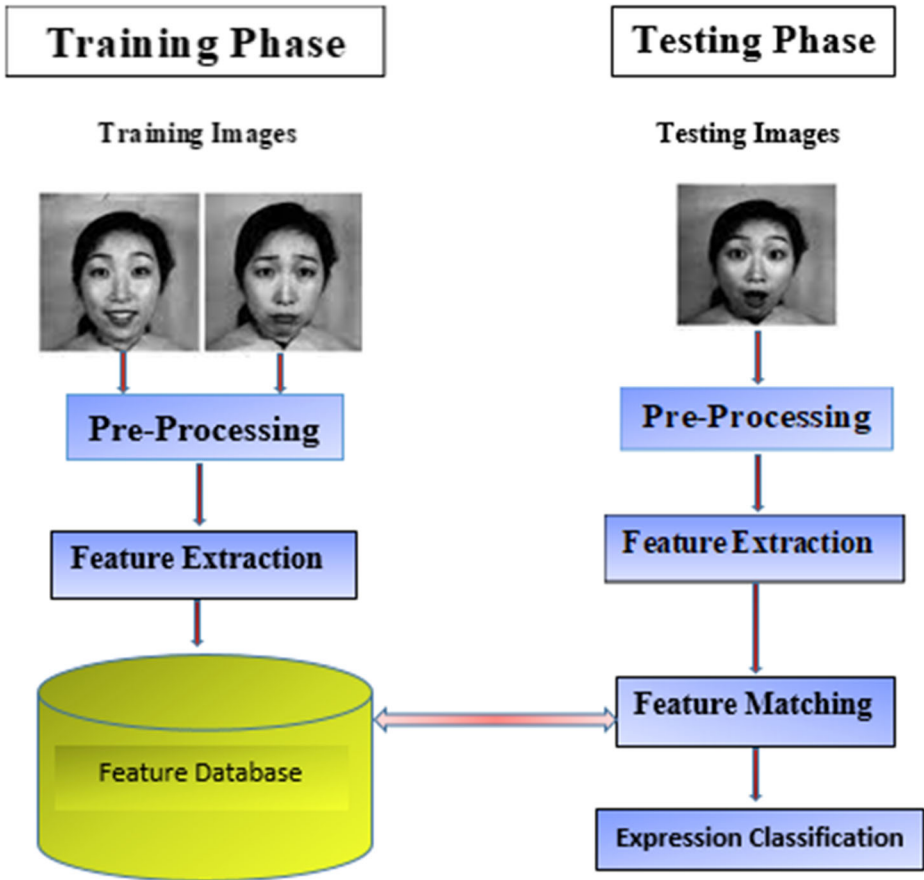


Fig. 5 Overall proposed approach Flow diagram

Calculating the degree of fuzziness is critical using a suitable objective function with an objective to minimize its value. The objective function is given in Eq. 1

$$J_{FCM} = (u, v) = \sum_{i=1}^c \sum_{k=1}^N \mu_{ik}^p \|x_k - v_i\|^2 \tag{1}$$

In Eq. 1,  $v = \{v_i\}_{i=1}^c$  are the cluster type,  $V = \{v_i\}_{i=1}^c$  is an array and is the position matrix,  $c$  represents the centre of cluster,  $N$  is the total number of pixels,  $x_k$  is the  $k^{th}$  pixel and  $v_i$  is the centroid of the  $i^{th}$  cluster. Further  $\|x_k - v_i\|^2$  is the distance between  $v_i$  (cluster centre) and  $x_k$  ( $k^{th}$  pixel)  $\mu_{ik}$  is the fuzzy membership value of  $k^{th}$  pixel in the  $i^{th}$  cluster. This membership value i.e.,  $\mu_{ik}$  has to satisfy the following constraint.

$$\mu_{ik} \in [0, 1], 1 \leq k \leq N, 0 < \sum_{k=1}^N \mu_{ik} < N, i \leq i \leq c, \text{ and } \sum_{i=1}^c \mu_{ik} = 1, 1 \leq k \leq N.$$

It is known from Eq. 1 that the objective function has minimized the distance between any pixel  $x_k$  and centroid of the  $i^{th}$  cluster  $v_i$ . Correspondingly,  $\mu_{ik}$ , which is fuzzy membership value of the  $k^{th}$  pixel associated with  $i^{th}$  cluster. Its value is in the range of [0–1] and its

association with all the clusters are interpreted from the value of the membership function. Based on the range of membership values, the following partition matrix is updated by equations given below in Eqs. (2) and (3).

$$\mu_{ik} = \frac{1}{\sum_{j=1}^c \left( d_{ik}^2 / d_{jk}^2 \right)^{1/p}} \quad (2)$$

$$v_i = \frac{\sum_{k=1}^N \mu_{ik}^p x_k}{\sum_{k=1}^N \mu_{ik}^p} \quad (3)$$

The FCM is applied for segmenting nose from the face and a sample result of segmentation is depicted in Figs. 3 and 4.

We have considered Yale dataset (Details are given in Experimental Section) for conducting experiments. The segmented nose region along with the original nose image are shown in Figure 2. It is observed from the Figure, that the FCM in the HSV color space can segment the nose region effectively for further processing and **analysis**.

### 3.2 Analysis of nose region using pyramid structure

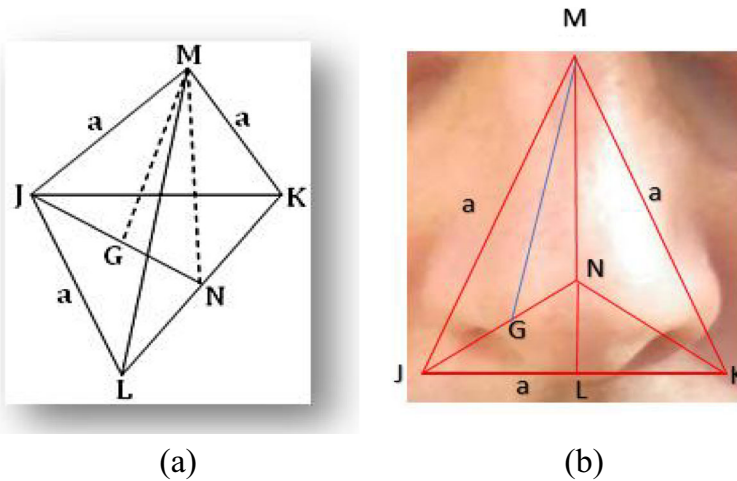
The procedure for extracting features along with required variables for the nose is depicted in Fig. 2. Most of the pyramids are characterized on their base. All the pyramid structure has a common property, that their sides are triangular shaped. While all the triangles are equilateral, the pyramid is termed a regular tetrahedron. In contrast, if the edges of the triangle have different lengths, the pyramid is termed an irregular tetrahedron. In the proposed approach, the properties of the triangular base pyramid are exploited to estimate expression/emotion on the human face. The nose on the face is represented as a triangular base pyramid and it is characterized as a regular or irregular tetrahedron based on various properties. The nose of the human face may not always be a regular tetrahedron and often goes with irregular. A sample tetrahedron is shown in Figure. 6

In Figs. 6(a) and (b), the Length of an edge is denoted as  $a$  and the Area is denoted as  $A$ . The area of the tetrahedron is the function of areas of three congruent equilateral triangles, which is given below

$$A = \left( x * \frac{(\sqrt{x})}{ya^2} \right) \quad (4)$$

In the above Eq. the value of  $x$  and  $y$  are 3 and 4 respectively and the unit of  $A$  is square units. The Area of the Whole surface ( $AW$ ) of the tetrahedron is the function of areas of four congruent equilateral triangles and is shown below.

$$AW = \left( x1 * \frac{(\sqrt{x})}{ya^2} \right) = \sqrt{y} * a^2 \text{ square units} \quad (5)$$



**Fig. 6** Pyramid and Tetrahedron model (a) Sample tetrahedron (b) Tetrahedron over lapped over a Nose

In the above Eq. the value of  $\times I$ ,  $x$  and  $y$  are 4, 3 and 4 respectively. The Volume ( $V$ ) of the regular tetrahedron is denoted as

$$V = \left( \frac{1}{3} \times \text{area of the base} \times \text{height} \right) \quad (6)$$

$$V = \left( \frac{1}{3} \right) \times \left( \frac{\sqrt{3}}{4} \right) \times a^2 \times \frac{(\sqrt{2})}{\sqrt{3}} \times a \quad (7)$$

$$V = \left( \frac{\sqrt{2}}{12} \right) \times a^3 \text{ cubic units} \quad (8)$$

The area ( $A$ ) and Area of Whole surface ( $AW$ ) given in Eqs. (4) and (5) are the function of length of edges of tetrahedron. Further, in the plane  $\Delta JLK$  we have,  $JN \perp LK$ , thus

$$JN^2 = (JL^2 - LN^2) = a^2 - \left( \frac{a}{2} \right)^2 = \frac{3a^2}{4}$$

$$\text{Now, } JG = \left( \frac{2}{3} \right) JN \text{ or } JG^2 = \left( \frac{4}{9} \right) \times JN^2$$

$$\text{or } JG^2 = \left( \left( \frac{4}{9} \right) \times \left( \frac{3}{4} \right) \times a^2 \right)$$

$$\text{or } JG^2 = \left( \frac{a^2}{3} \right)$$

In addition,  $MG \perp JG$  and  $JM = a$  and thus, from  $\Delta JGM$  we get,

$$MG^2 = JM^2 - JG^2$$

$$\text{or } MG^2 = \left( a^2 - \left( \frac{a^2}{3} \right) \right)$$

$$\text{or } MG^2 = \left( \frac{2a^2}{3} \right)$$

As a result,

$$MG = \frac{\sqrt{2}a}{\sqrt{3}} = \text{height of the regular tetrahedron} \quad (9)$$

The height of regular tetrahedron given in above Eq. 9 is the function of  $a$ . As a result, the tetrahedron based model of the nose of human face can be effectively used for estimating emotion. For an emotion, say, happy, the deformation on the nose area is measured and compared with neutral nose. A similarity measure, say, Manhattan distance is used for calculating the deformation from which the emotional type is estimated.

### 3.3 Classification of facial expressions

In this subsection, the procedure of modelling feature vectors such as area, area of whole surface, Volume, Height, and the points on the nose from  $a$  to  $m$  is presented. These features are later given as input to classifiers. Support Vector Machine, Multi-Layer Perceptron, and Random Forest classifiers are used for classifying facial expressions. The section below presents the procedure of classification by these classifiers.

#### 3.3.1 Support vector machine (SVM)

SVM is the proposed approach that has been used for classification as it is one of the well-known machine learning algorithms. SVM has a set of hyperplanes in an infinite dimensional space and it addresses a binary pattern classification problem. As a result, it is considered as the linear and nonlinear SVM. In SVM, the margin between various classes is minimized effectively while maintaining a minimum training set error. It is possible to find a global minimum very easily as the SVM is a convex quadratic programming method. However, it is important to set and tune various important parameters for the SVM classifier. Some of the important parameters are penalty and smoothness parameters of the radial-based function. The feature database is trained by the SVM as given below. The training dataset  $TS$  is represented as follows.

$$TS = \{(fv_1), (fv_2), (fv_3) \cdots (fv_n)\}, c_j \in \{\text{Emotional Expressions}\}, (fv_i) \in FV^d \quad (10)$$

In the above Eq. (10),  $fv_i$  represents the feature vector extracted from the nose,  $c_j$  represents the corresponding expression category and  $n$  is the size of the dataset. The feature vector is  $n$ -fold cross validated for training the SVM. To train the multi-class SVM classifier for all the well-known six emotional expressions, the images from CK++ and JAFFE datasets are grouped based on their emotional category. The training and testing set of the data set is presented to SVM for training the classifier as it is supervised classifier. As a result, the classifier effectively finds the decision boundary between the various emotional categories as shown in Fig. 7. During the training, linear kernel is used such that each input  $fv_i$  has  $d$  attributes to denote its class  $c_j = \{\text{Emotional Expressions}\}$ . The training dataset  $TS$  is shown in Eq. (10).

#### 3.3.2 Multi-layer perceptron (MLP)

MLP is considered as a feed forward neural network. Generally, it consists of input, hidden and output layers. The number of hidden layers of the network can be one or more depending

on the problem. The data is presented to the input layer, processed by both input and hidden layers and fed forward to the output layer as represented in Fig 7. The MLP uses supervised learning techniques called back propagation for training. MLP solves problems which are not linearly separable. For instance, let us consider feature vector from all regions of the face. If we have  $m$  input data  $\{(fv_1), (fv_2), (fv_3), \dots, (fv_m)\}$ , it is called as  $m$  features. A feature, say one row of the feature database is considered as one variable, which influences a specific outcome of emotional category. Each of the  $m$  features is multiplied with a weight ( $w_1, w_2, \dots, w_m$ ), sum them all together for obtaining a dot product as given below.

$$W(FV) = w_1(fv_1) + w_2(fv_2) + \dots + w_m(fv_m) \quad (11)$$

$$W(FV) = \sum_{i=1}^m w_i(fv_i) \quad (12)$$

Here,  $m$  features are given as input in the form of  $(FV)$  to the input layers of the MLP as shown in Fig. 8. Suitable weights are determined during the training process such that the dot product is calculated between weights and inputs as shown in Eqs. (11) and (12). Similarly, the weights for each neuron in the hidden layers are also calculated. The dot product between the weights and input values of hidden layer is similar to the input layer and as a result, the hidden layer is considered as yet another single layer perceptron.

### 3.3.3 Random Forest

In random forest classifier, more number of individual decision trees are present and operates as ensemble. The individual decision trees predict a class the class which got the most votes is considered as predicted class. A typical structure of Random Forest is shown Fig. 9.

The procedure of training the Random Forest in the proposed approach is as follows. We considered the feature vector ( $P$ ), extracted from each data sets and used for training the classifier.

$$P = p_1, p_2, \dots, p_n \quad (13)$$

$$C = c_1, c_2, \dots, c_m \quad (14)$$

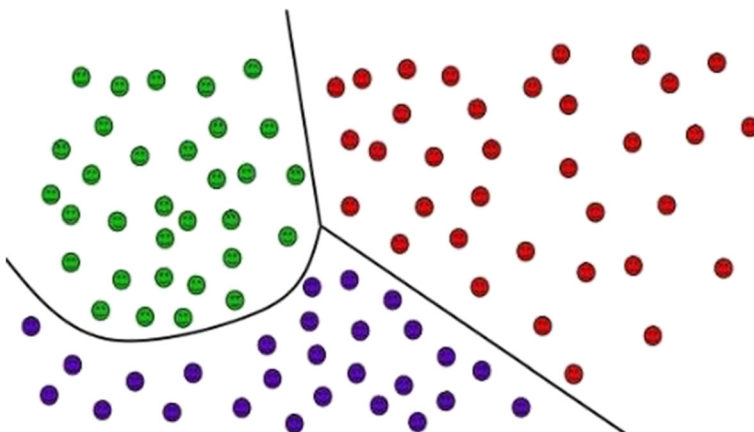


Fig. 7 Sample multiclass output space of SVM

In the above Eqs. (13-14),  $n$  is the number of data samples for training and  $m$  is number of emotional expression classes. Now, given the training set  $P$ , with responses  $Q$ , select a random sample to replace the training set by repeatedly performing bagging for  $B$  times. Once the training is over, the predictions on unseen samples  $p'$  can be estimated by averaging the predictions from all the individual regression tree on  $p'$ , which is given below Eq. (15).

$$\hat{f} = \frac{1}{B} = \sum_{b=1}^B f_b(p') \tag{15}$$

The proposed approach minimizes the variance of the model given above to have strongly de-correlating trees. In addition, the uncertainty of prediction on sample  $p'$  can be written as follows.

$$\sigma = \sqrt{\frac{\sum_{b=1}^B (f_b(p') - \hat{f})^2}{B-1}} \tag{16}$$

In the above Eq. (16),  $B$  is the number of samples.

In all the classifiers, the training is n-fold cross validation to ensure the participation of each sample in both training and testing for improving the classification accuracy. The performance of each classifier while conducting the experiments is presented in Section 4.

## 4 Experimental results

### 4.1 Details of dataset

In this section, the performance of the proposed geometrical feature of the nose using the properties of Pyramid is presented. We have considered the well-known benchmark datasets such as Japanese Female Facial Expression (JAFPE), Cohn–Kanade Facial Expression (CK++), and Indian Spontaneous Expression Database (ISED) for experiments. The JAFE database contains 213 images and seven expression classes namely *Sadness*, *Neutral*, *Happiness*, *Disgust*, *Anger*, *Fear*, and *Surprise* for 10 different Japanese models. All the face images

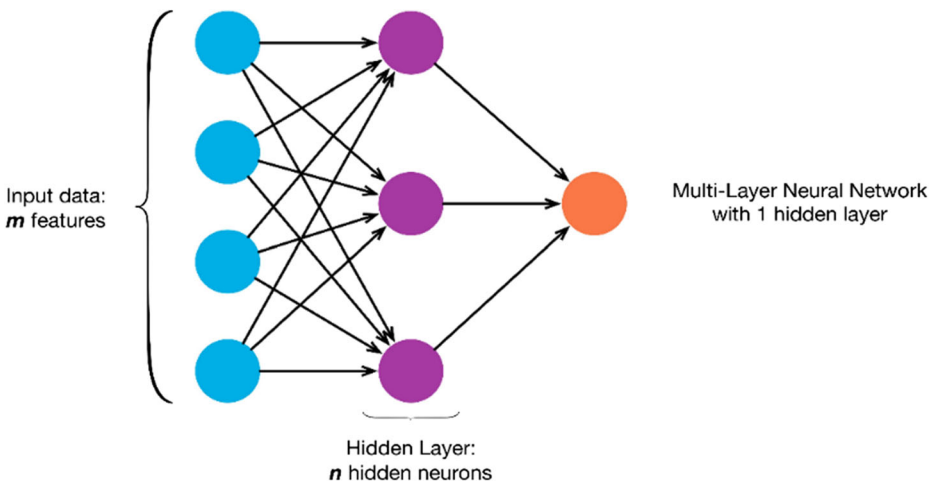


Fig. 8 Typical Structure of Multi-Layer Perceptron

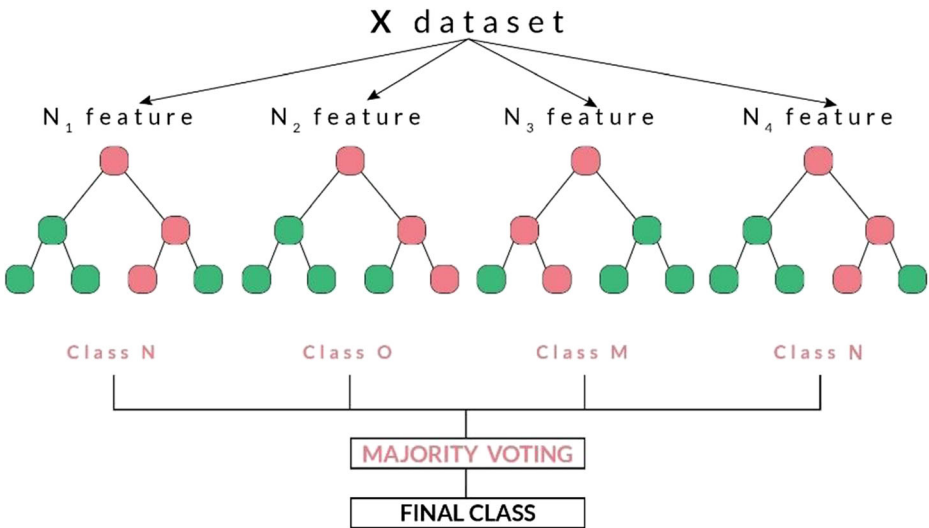


Fig. 9 Random Forest

are resized to  $256 \times 256$ . The Cohn–Kanade Facial Expression database (CK++) is developed by Carneo Milan University Pittsburgh and has 593 images belonging to 123 different subjects. Happy et al. [25] has created ISED dataset for analyzing facial expression. This dataset consists of images with spontaneous expressions of both male and female genders from India as birthplace. The dataset consists of images of 50 persons with 428 video frames. Fan et al. [17] have used Kolmogorov-Smirnov Predictive Accuracy (KSPA) and Augmented Dickey-Fuller (ADF) for testing the data. The data is on fluctuation in electricity load, which is a time series data and thus, Kolmogorov-Smirnov Predictive Accuracy (KSPA) and Augmented Dickey-Fuller (ADF) type methods can be used to test the result. However, in this work, we have static data and Kolmogorov-Smirnov Predictive Accuracy (KSPA) and Augmented Dickey-Fuller (ADF) may not be used for testing. The effect of illumination effects is

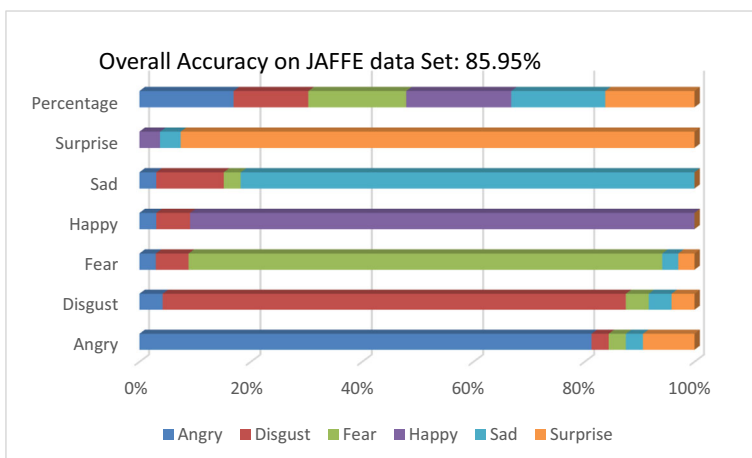
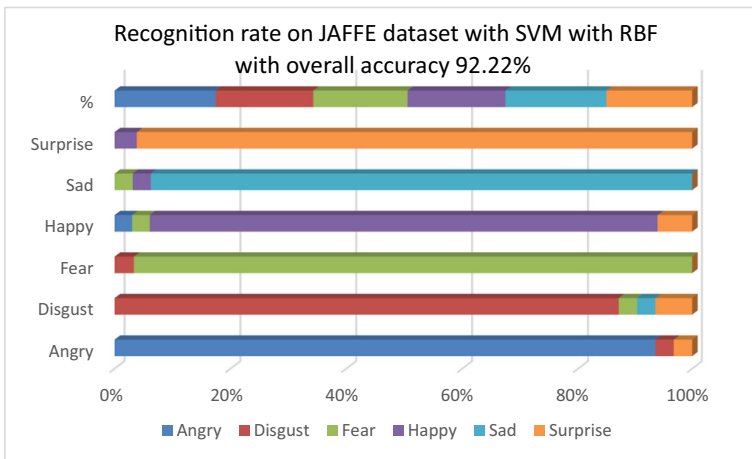


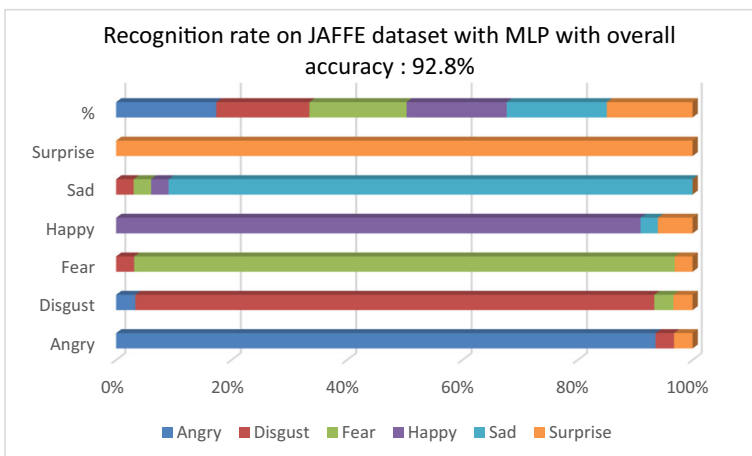
Fig. 10 Recognition rate on Jaffe Dataset





**Fig. 11** Recognition rate on JAFFE dataset with SVM and RBF

handled using Histogram equalization. The experiment is conducted on system with 11th Gen Intel(R) Core(TM) i5-1135G7 @ 2.40 GHz, 2401 MHz, 4 Core(s), 8 Logical Processor(s). The Operating System is Microsoft Windows 11 Home Single Language with 8 GB RAM. We have used the Open CV library to segment the face portion from the whole image using the frontal face detector. The illumination effect is handled by histogram equalization such that the effect of illumination is normalized. The Simulink to the interface of open CV for MATLAB 2013 environment is used for experiments. The experiments are conducted to estimate the well-known expressions such as neutral, surprise, disgust, happy, feat, sad, and anger. We have used the recognition rate / classification accuracy as a performance measure for each emotional category and evaluated each dataset. The evaluation of the JAFFE dataset for various expression categories is presented in Fig. 10 and the overall accuracy is 85.95%. Figure 10 is the same as the confusion matrix.



**Fig. 12** Recognition rate on JAFFE dataset with MLP

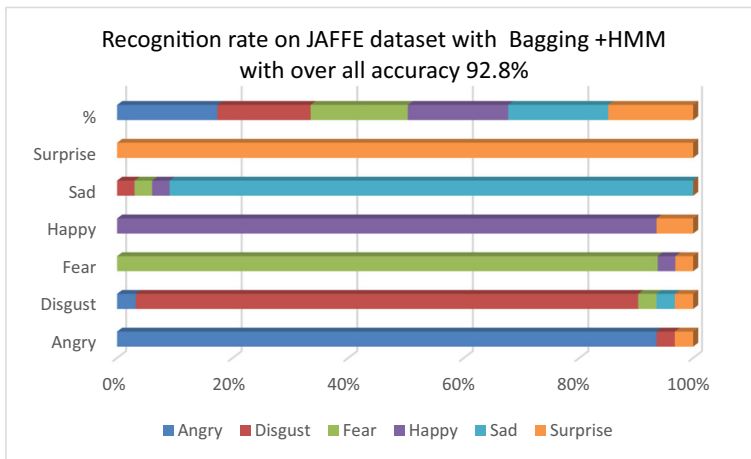


Fig. 13 Recognition rate on JAFFE dataset with Bagging +HMM

#### 4.2 Performance of the proposed approach using machine learning algorithms

It is known in Fig. 8 that the recognition rate/classification accuracy is 85.95% on JAFFE dataset. This is found to be on the lower side and has to be improved. Thus, in this paper, we have used Random Forest (RF), Multi-Layer Perceptron (MLP), and Support Vector Machine (SVM) for improving performance. It is known that there are set of base parameters for each algorithm and all these parameters are fixed based on experiments. We have used LibSVM library for implementing SVM and used Linear, Polynomial, Radial Basis Function (RBF) kernels. The recognition rate of SVM with RBF is shown in Fig. 9. During the training stage, the C and Gamma in RBF are used as 20 and 0.1 respectively. The recognition rate on JAFFE dataset using Multi-Layer Perceptron is shown in Fig. 10. The configuration of MLP is: input layer is four neurons, hidden layer consists of twenty neurons and output layer consists of seven neurons, the learning rate is 0.3, momentum is 0.2, and training time is 500. Similarly, we have used Ensemble learning is used with Ad boost Meta - algorithm along with the random forest. The training model is built

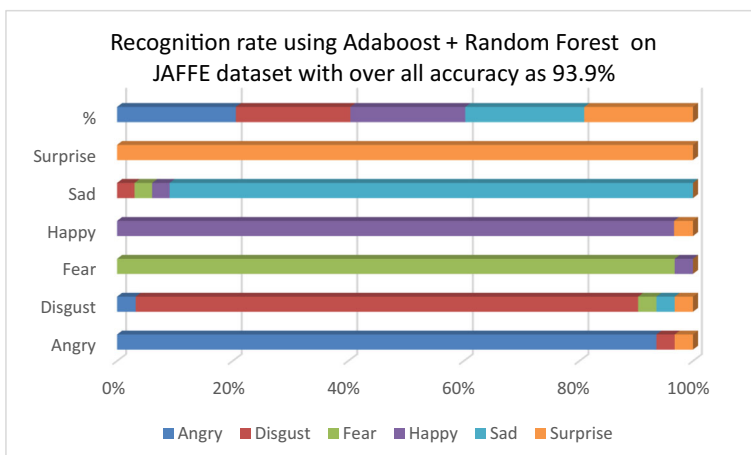


Fig. 14 Recognition rate on JAFFE dataset using Ad boost + Random Forest

**Table 1** Performance Comparison - the recognition Rate for various Emotions

| State-of-the-art approaches | Angry        | Disgust     | Fear        | Happy      | Sad          | Surprise    | Recognition Rate (%) |
|-----------------------------|--------------|-------------|-------------|------------|--------------|-------------|----------------------|
| Proposed Work               | <b>98.00</b> | <b>94.6</b> | <b>94.9</b> | <b>100</b> | <b>98.00</b> | <b>96.7</b> | <b>97.1</b>          |
| Happy and Routray [24]      | 100.0        | 86.2        | 93.8        | 96.8       | 77.4         | 96.7        | 91.8                 |
| Alhussein [1]               | 83.3         | 68.1        | 67.2        | 98.1       | 78.4         | 88.3        | 80.5                 |
| Li et al. [29]              | 94.4         | 88.6        | 89.0        | 96.0       | 92.2         | 96.5        | 92.7                 |
| Zeng et al. [50]            | 93.3         | 86.2        | 89.0        | 100        | 93.3         | 95.3        | 92.8                 |
| Bilkhu et al. [6]           | 92.4         | 90.8        | 89.5        | 94.2       | 84.2         | 88.3        | 89.9                 |
| Shahid et al. [34]          | 96.7         | 90.1        | 93.8        | 95.5       | 90.3         | 93.3        | 93.2                 |
| Yao et al. [47]             | 96.7         | 88.0        | 91.8        | 95.6       | 93.6         | 90.0        | 92.6                 |
| Ghazouani [21]              | 96.7         | 90.0        | 92.8        | 95.6       | 93.6         | 90.0        | 93.1                 |
| Srinivasu et al. [39]       | 66.8         | 84.5        | 67.6        | 86.3       | 86.2         | 79.3        | 81.1                 |

for this experiment has used the bagging with Hidden Markov Model. The random subset of the dataset is generated by applying re-sampling for the random forest tree algorithm. The testing and trading phase of all the classifiers have used 10-fold Cross-Validation to avoid over and underfitting. The recognition rate of various classifiers on JAFFE dataset is presented in Figs. 11, 12, 13 and 14. The overall accuracy from the Fig. 11 is 92.22%

It is found that the recognition rate of most of the classification algorithms is from 95% to 96%. It is found that the performance of the proposed approach is good for angry and sad, happy and disgust is moderate and low for a surprise. Also, in the JAFFE data set, surprise images are also classified as normal. This is due to the fact that these normal images with an open (close) mouth is having a similar feature set as a surprise. Surprise images in JAFFE are normal with a close or slightly open mouth. Based on the results, the proposed approach has performed well on surprise expression having random forest with ensemble classifier configuration as shown in Fig. 14.

### 4.3 Performance comparison

In Table 1, we present the comparative analysis of the results and JAFFE dataset is considered as data set. It is observed from the results that the proposed approach performs well for the expression categories of *happy*, *angry* and *surprise*.

**Table 2** Comparisons of recognition rates of various approaches for different Emotions on CK+ Database

| State-of-the-art approaches | Angry       | Disgust      | Fear        | Happy      | Sad         | Surprise   | Recognition Rate (%) |
|-----------------------------|-------------|--------------|-------------|------------|-------------|------------|----------------------|
| <b>Proposed Work</b>        | <b>98.1</b> | <b>95.00</b> | <b>95.5</b> | <b>100</b> | <b>95.7</b> | <b>100</b> | <b>97.3</b>          |
| Gu et al. [23]              | 63.1        | 91.9         | 84.8        | 89.8       | 86.4        | 97.1       | 85.5                 |
| Zhang et al. [52]           | 83          | 82           | 65          | 90         | 60          | 83         | 77.1                 |
| Happy and Routray [24]      | 94.3        | 87.8         | 91.3        | 96.2       | 95.4        | 98.5       | 93.9                 |
| Alhussein [1]               | 96.4        | 92.7         | 93.1        | 99.6       | 94.5        | 99.7       | 96                   |
| Li et al. [29]              | 94.4        | 71.3         | 80.3        | 94.4       | 87.2        | 98.4       | 87.6                 |
| Ghimire et al. [22]         | 87.0        | 91.9         | 91          | 96.9       | 84.6        | 91.2       | 90.4                 |
| Zeng et al. [50]            | 95.2        | 85.2         | 83.3        | 97.1       | 90.2        | 98.3       | 91.5                 |
| Bilkhu et al. [6]           | 94.6        | 74.6         | 68.6        | 97.7       | 88.9        | 92.5       | 86.1                 |
| Yao et al. [47]             | 95.3        | 93.8         | 90.5        | 98.5       | 88.6        | 96.1       | 93.8                 |
| Shahid et al. [34]          | 97.5        | 94.6         | 82.6        | 98.2       | 92.3        | 100.0      | 94.2                 |
| Ghazouani [21]              | 97.5        | 94.7         | 96          | 100.0      | 96.7        | 98.0       | 97.1                 |
| Srinivasu et al. [39]       | 94.3        | 83.7         | 76.0        | 96.1       | 74.0        | 92.6       | 86.1                 |

The recognition rate is presented in Table 1 and the values are highlighted to emphasize the performance enhancement. The reason for performance enhancement is due to the fact that the proposed approach has captured the features of the nose effectively. The deformation due to various expressions in the nose has reflected at various points of the pyramid and thus has achieved good results. In the results, the lowest recognition rate is achieved by Alhussein [1], which is 80.5%. The methods proposed by Shahid et al. [34] and Ghazouani [21] have achieved a 93% recognition rate. It is also observed that the performance of methods proposed by Yao et al. [47] and Li et al. [29] has achieved 92% accuracy. However, overall, the performance of the proposed nose feature using a pyramid structure is good. While observing the performance of Srinivasu et al. [39], it is 81.1% and the reason is that the segmentation algorithm works well in the medical images.

The recognition rate/classification accuracy of the proposed approach on the CK++ dataset is shown in Table 2. It is observed that the recognition rate of *anger*, *surprise*, and *happiness* is notable compared to other expressions. The recognition rate of *fear*, *disgust*, and *sad* conflicts. Yao et al. [47] have achieved 93.8% as the recognition rate and Shahid et al. [34] have achieved 90.2%. In Ghazouani [21], *anger* has contraction with *surprise*. The recognition rate of the nose feature is good for *happiness*, *surprise*, and *anger* not encouraging for other expressions such as *sad*, *disgust*, and *fear*. It is observed from Table. 2 that the recognition rate of the proposed approach for the nose feature is (97.3%), which is encouraging compared to others in the literature.

## 5 Conclusion

In this section, we conclude the paper with our contributions, limitations, and future perspective of the study. We have proposed a geometrical model using the concept of pyramid and tetrahedron. The properties of these are effectively used for modelling the nose component of the human face for estimating the expression and deriving the emotion. The feature vector from the nose is extracted, superimposed on the neutral face, and compared with the images in the database. The degree of deformation is used for estimating the expression and deriving the emotion. The classification accuracy is used as a performance metric and we have used JAFE and CK++ datasets for experiments. The performance of the proposed approach is compared with the contemporary methods. It was found that the performance is encouraging. However, one of the limitations of the proposed approach is that it captures only the micro-level deformation of the nose. Thus, this method may not be suitable for any real-time application, which uses macro-level deformation as a feature. This issue can be handled in future by capturing the features from all the components of the face. Thus, the Marco level feature can be constructed for better performance.

**Funding** We have not received funding from any funding agency for the work presented in this paper.

## Declarations

**Conflict of interest** We have no conflicts of interests and competing interests in the presented work.

## References

1. Alhussein M (2016) Automatic facial emotion recognition using weber local descriptor for e-Healthcare system. *Cluster Comput-J Netw Softw Tools Appl* 19(1):99–108

2. Avani VS, Shaila SG, Vadivel A (2021a) Geometrical features of lips using the properties of parabola for recognizing facial expression. *Cogn Neurodyn* 15:481–499. <https://doi.org/10.1007/s11571-020-09638-x>
3. Avani VS, Shaila SG, Vadivel A (2021b) Interval graph of facial regions with common intersection salient points for identifying and classifying facial expression. *Multimed Tools Appl* 80:1–24. <https://doi.org/10.1007/s11042-020-09806-5>
4. Benitez-Quiroz CF, Srinivasan R, Martinez AM (2016) EmotioNet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild, in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit*, pp. 5562–5570
5. Bezdek JC (1973) *Fuzzy mathematics in pattern classification*, Ph.D. dissertation, Appl. Math. Center, Cornell Univ., Ithaca, NY
6. Bilkhu MS, Gupta S, Srivastava VK (2019) Emotion classification from facial expressions using cascaded regression trees and SVM, in *Computational Intelligence: Theories, Applications and Future Directions-Volume II*, Springer, pp. 585–594
7. Biswas S (2015) An efficient expression recognition method using Contourlet transform. *Int Conf Percept Mach Intell* pp 167–174
8. Biswas B, Mukherjee AK, Konar A (1995) Matching of digital images using fuzzy logic. *AMSE Publication* 35(2):7–11
9. Bolotnikova A, Demirel H, Anbarjafari G (2017) Real-time ensemble based face recognition system for NAO humanoids using local binary pattern. *Analog Integr Circuits Signal Process* 92(3):467–475
10. Calvo MG, Fernández-Martín A, Nummenmaa L (2014) Facial expression recognition in peripheral versus central vision: Role of the eyes and the mouth. *Psychol Res* 78(2):180–195
11. Chakraborty A, Konar A, Chakraborty UK, Chatterjee A (2009) Emotion recognition from facial expressions and its control using fuzzy logic. *IEEE Trans Syst Man Cybern A Syst Humans* 39(4):726–743
12. Cohen I (2000) Facial expression recognition from video sequences, M.S. thesis, Univ. Illinois Urbana-Champaign, Dept. Elect. Eng., Urbana, IL
13. Demir Y (2014) A new facial expression recognition based on curvelet transform and online sequential extreme learning machine initialized with spherical clustering. *Neural Comput Appl* 27:131–142. <https://doi.org/10.1007/s00521-014-1569-1>
14. Drume D, Jalal AS (2012) A multi-level classification approach for facial emotion recognition, in *Proc. Int Conf Comput Intell Comput Res.*, Coimbatore, India, , pp. 288–292
15. Ekman P, Friesen WV (1975) *Unmasking the face: a guide to recognizing emotions from facial clues*. Prentice-Hall, Englewood Cliffs
16. Ekman P, Friesen W (1978) *Facial action coding system: a technique for the measurement of facial movement*. Consulting Psychologists, Palo Alto, CA, USA, Tech. Rep.
17. Fan et al (2021) Forecasting short-term electricity load using hybrid support vector regression with grey catastrophe and random forest modeling. *Util Policy* 73(2021):101294
18. Fellenz WA, Taylor JG, Cowie R, Douglas-Cowie E, Piat F, Kollias S, Orovas C, Apolloni B (2000) On emotion recognition of faces and of speech using neural networks, fuzzy logic and the ASSESS systems, in *Proc. IEEE-INNS-ENNS Int. Joint Conf. Neural Netw*, pp. 93–98
19. Florea C, Florea L, Badea MA, Vertan C (2019) Annealed label transfer for face expression recognition. In *British Machine Vision Conference (BMVC)*
20. Gao Y, Leung MKH, Hui SC, Tananda MW (2003) Facial expression recognition from line-based caricatures. *IEEE Trans Syst, Man, Cybern A, Syst, Humans* 33(3):407–412
21. Ghazouani H (2021) A genetic programming based feature selection and fusion for facial expression recognition. *Appl Soft Comput* 103:107173
22. Ghimire D, Lee J, Li Z-N, Jeong S (2017) Recognition of facial expressions based on salient geometric features and support vector machines. *Int J Multimed Tools Appl* 76:7921–7946
23. Gu W, Xiang C, Venkatesh YV, Huang D, Lin H (2012) Facial expression recognition using radial encoding of local Gabor features and classifier synthesis. *Pattern Recognit* 45(1):80–91
24. Happy SL, Routray A (2015) Automatic facial expression recognition using features of salient facial patches. *IEEE Trans Affective Comput* 6(1):1–12
25. Happy SL, Routray A (2017) Fuzzy histogram of optical flow orientations for micro-expression recognition. *IEEE Trans Affect Comput* 10(3):394–406
26. Izumitani K, Mikami T, Inoue K (1984) A model of expression grade for face graphs using fuzzy integral. *Syst Control* 28(10):590–596
27. Kumar S, Bhuyan MK, Chakraborty BK (2016) Extraction of informative regions of a face for facial expression recognition. *IET Comput Vis* 10:567–576. <https://doi.org/10.1049/iet-cvi.2015.0273>

28. Li X, Ji Q (2005) Active affective state detection and user assistance with dynamic Bayesian networks. *IEEE Trans Syst, Man, Cybern A, Syst, Humans* 35(1):93–105
29. Li S, Deng W, Du J (2017) Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild, in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 2584–2593
30. Liu Y, et al. (2016) Facial expression recognition with PCA and LBP features extracting from active facial patches, 2016 IEEE International Conference on Real-time Computing and Robotics (RCAR), 2016, pp. 368–373, <https://doi.org/10.1109/RCAR.2016.7784056>.
31. Mahersia H, Hamrouni K (2015) Using multiple steerable filters and Bayesian regularization for facial expression recognition. *Eng Appl Artif Intell* 38:190–202. <https://doi.org/10.1016/j.engappai.2014.11.002>
32. Minaee S, Minaei M, Abdolrashidi A (2021) Deep-emotion: facial expression recognition using attentional convolutional network. *Sensors* 21:3046
33. Salmam FZ, Madani A, Kissi M (2016) Facial expression recognition using decision trees. *IEEE. In: 13th Int Conf Comput. Graph Imaging Vis*, pp 125–130. <https://doi.org/10.1109/CGiV.2016.33>
34. Shahid AR, Khan S, Yan H (2020) Contour and region harmonic features for sub-local facial expression recognition. *J Vis Commun Image Represent* 73:102949
35. Siddiqi M, Ali R, Khan A, Park YT, Lee S (2015) Human facial expression recognition using stepwise linear discriminant analysis and hidden conditional random fields. *IEEE Trans Image Process* 24:1386–1398. <https://doi.org/10.1109/TIP.2015.2405346>
36. SivaSai JG, Srinivasu PN, Sindhuri MN, Rohitha K, Deepika S (2021) An automated segmentation of brain mr image through fuzzy recurrent neural network. In: Bhoi A, Mallick P, Liu CM, Balas V (eds) *Bio-inspired Neurocomputing. Studies in Computational Intelligence*, vol 903. Springer, Singapore. [https://doi.org/10.1007/978-981-15-5495-7\\_9](https://doi.org/10.1007/978-981-15-5495-7_9)
37. Srinivasu NP et al (2020a) A comparative review of optimization techniques in segmentation of brain MR images. *J Intell Fuzzy Syst* 38(5):6031–6043. <https://doi.org/10.3233/JIFS-179688>
38. Srinivasu NP et al (2020b) A comparative review of optimisation techniques in segmentation of brain MR images. *J Intell Fuzzy Systems* 38(5):6031–6043. <https://doi.org/10.3233/JIFS-179688>
39. Srinivasu N et al (2021) An AW-HARIS based automated segmentation of human liver using CT images. *69(3):3303–3319*. <https://doi.org/10.32604/cmc.2021.018472>
40. Stockman G, Shapiro LG (2001) *Computer vision*. Prentice Hall PTR, USA. isbn:978-0-13-030796-5
41. Swaminathan A, Vadivel A, Arock M (2020) FERCE: facial expression recognition for combined emotions using FERCE algorithm. *IETE J Res*. <https://doi.org/10.1080/03772063.2020.1756471>
42. Uwechue OA, Pandya SA (1997) *Human face recognition using third-order synthetic neural networks*. Kluwer, Boston
43. Vadivel A, Sural S, Majumdar AK (2009) Robust histogram generation from the HSV space based on visual colour perception. *Int J Signal Imaging Syst Eng* 1(3–4):245–254
44. Vivek TV, Reddy GRM (2015) A hybrid bioinspired algorithm for facial emotion recognition using CSO-GA-PSO-SVM, in *Proc. 5th Int. Conf. Commun. Syst. Netw. Technol., Gwalior, India*, pp. 472–477
45. Wang K, Peng X, Yang J, Lu S, Qiao Y (2020a) Suppressing uncertainties for large-scale facial expression recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*:6896–6905
46. Wang K, Peng X, Yang J, Meng D, Yu Q (2020b) Region attention networks for pose and occlusion robust facial expression recognition. *IEEE Trans Image Process* 29:4057–4069
47. Yao L, Wan Y, Ni H, Xu B (2021) Action unit classification for facial expression recognition using active learning and SVM. *Multimed Tools Appl* 80:1–15
48. You D, Hamsici OC, Martinez AM (2011) Kernel optimization in discriminant analysis. *IEEE Trans Pattern Anal Mach Intell* 33(3):631–638
49. Zavarez MV, Berriel RF, Oliveira-Santos T (2017) Cross-database facial expression recognition based on fine-tuned deep convolutional network. In *Proceedings of the 2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, Niteroi, Brazil, 17–20; pp. 405–412
50. Zeng J, Shan S, Chen X (2018) Facial expression recognition with inconsistently annotated datasets. In *European Conference on Computer Vision (ECCV)*, pages 227–243, Cham
51. Zhang H, Liu D, Wang Z (2009) *Controlling chaos: suppression, synchronization and Chaotification*, New York: Springer-Verlag
52. Zhang L, Jiang M, Farid D, Hossain MA (2013) Intelligent facial emotion recognition and semantic-based topic detection for a humanoid robot. *Expert Syst Appl* 40:5160–5168
53. Zhang YD, Yang ZJ, Lu HM, Zhou XX, Phillips P, Liu QM, Wang SH (2016) Facial emotion recognition based on biorthogonal wavelet entropy, fuzzy support vector machine, and stratified cross validation. *IEEE Access* 4:8375–8385

54. Zhao H, Wang Z (2008) Facial action units recognition based on fuzzy kernel clustering. in Proc. 5th Int. Conf. Fuzzy Syst. Knowl. Discov, pp. 168–172
55. Zhong F, Yan S, Liu L, Liu K (2018) An effective face recognition framework with subspace learning based on local texture patterns, in 2018 5th International Conference on Systems and Informatics (ICSAI), pp. 266–271

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.