



# Facial Emotion Intensity: A Fusion Way

Ankita Pandey<sup>1</sup> · Arun Kumar<sup>1</sup>

Received: 27 September 2021 / Accepted: 1 February 2022 / Published online: 13 February 2022  
© The Author(s), under exclusive licence to Springer Nature Singapore Pte Ltd 2022

## Abstract

Many scientific works have been conducted for developing the Emotion intensity recognition system. But developing a system that is capable to estimate small to peak intensity levels with less complexity is still challenging. Therefore, we propose an effective facial emotion intensity classifier by fusion of the pre-trained deep architecture and fuzzy inference system. The pre-trained architecture VGG16 is used for basic emotion classification and it predicts emotion class with the class index value. By class index value, images are sent to the corresponding Fuzzy inference system for estimating the intensity level of detected emotion. This fusion model effectively identifies the facial emotions (happy, sad, surprise, and angry) and also predict the 13 categories of emotion intensity. This fusion model got 83% accuracy on a combined dataset (FER 2013, CK+ and KDEF). The performance and findings of this proposed work are further compared with state-of-the-art models.

**Keywords** Pre-trained VGG · Fuzzy inference system · Basic emotion recognition · Emotion intensity estimation · Computer vision

## Introduction

Emotion intensity has attracted researchers for many years. Emotions are a complex psychological state that reveals feelings, moods, thoughts, and reactions while the intensity of emotion is a non-monotonic function that evaluates the strength level of emotion.

Researchers explore the depth of emotions to tackle the problem in their field. We can estimate its versatility by marking its applications in many fields including medical science, psychology, computer science, security system, recommender system, education, marketing, social science, human science, and many more. For instance, in medical science depression counseling [1], Autism spectrum disorders [2], and for patients seeking therapies [3] emotions are study. In psychology, many theories are proposed by researchers to illustrate the origin of emotions, its neurobiological explanation, along many aspects of emotions. Even

to enhance the power of security systems, emotion and its intensity are examined by developers [4, 5]. However, emotions play a crucial role in education as suggested by leading researchers [6]. Many e-learning processes depict its vital importance. Furthermore, to study mental processes and disorders, human science uses emotion. Emotion in marketing is termed emotional marketing. It acts as a communication between buyer and seller to reveal the experience of the buyer with a particular product. In other studies, smile intensity linked to life satisfaction and years lived. In computer science, the study of emotions is called affective computing and it is based on several features like speech [7], text [8], facial expression [9], EEG signal [10], body posture [11] and context [12].

In the domain of Computer vision, a lot of work has been done in facial expression recognition (FER) field. Recently, a survey held by the researcher [13] on the FER system provided a complete summary of recent works based on deep learning. Some of FER works use CNN-based architecture and other Fuzzy inference systems (FIS) to address the problem. Besides these, another popular technique is transfer learning. The reason behind their popularity is they reduce training time and increase performance. For instance Sajjanhar et al. [9] proposed a FER system based on transfer learning. They evaluated performance of FER task through pre-trained models Inception-v3, VGG, and VGG face.

---

✉ Ankita Pandey  
ankitapandey0595@gmail.com

Arun Kumar  
Arun\_pal1969@yahoo.co.in

<sup>1</sup> Department of Mathematics, Statistics and Computer Science, G. B. Pant University of Agriculture and Technology, Pantnagar 263145, India

Rassadin et al. [14] used transfer learning for face identification in their group-level emotion recognition work. They extracted landmark points by Dlib and calculated pair-wise distance between them to learn face features. Then they applied classifiers such as logistic regression, support vector regression, gradient boosting tree, and random forest. A recent study [15] have shown that among VGG16, ResNet152V3, Inception V3, and Xception, VGG16 was superior for FER on a combined dataset of CK+ and JAFFE with an accuracy of 83.16%. Many geometrical-based researches [16–18] are relying on facial landmark points for finding facial emotions. Recently, Amal et al. [19] did a real-time emotion recognition work on FER 2013 dataset using local binary patterns (LBP) for face detection, Dlib for landmark points extraction, and constructed a CNN with histogram of oriented gradients (HOG) features. This work got 75.1% accuracy shows that higher classification accuracy for real-time FER systems is still challenging.

On the other hand, Nicolai and Choi [20] introduced a fuzzy-based FER system with an accuracy of 78.8% on the JAFFE dataset. Also, Farahani et al. [21] considered a Mamdani-based fuzzy system with face features eye (opening, width ratio) and mouth (opening, width) for FER then got 78.8% accuracy. Another work with a similar concept is Chakraborty et al. [22]. The next FER system [16] was based on fuzzy logic for 5 basic emotions except neutral and used the concept of finding displacement of 17 landmark points between expressive and neutral frames. They evaluated their performance on the CK+ dataset. Again Bahreini et al. [17] presented software by calculating 54 cosine values from the face with 37 FURIA fuzzy rules they predicted 6 basic emotions with 83.2% accuracy.

Both CNN and FIS systems also proposed emotion intensity-related work, like Witzig et al. [23] presented smile intensity work base on the combined structure of CNN and RNN. Esau et al. [18] developed a real-time fuzzy-based emotion intensity work for four emotions happy, sad, and angry, and fear. In this work, facial features were represented by 6 angles and achieved a recognition accuracy of 72%. Vinola and Vimala Devi [24] generated fuzzy-based smile intensity work and considered the Euclidean distance concept between landmark points. This work achieved a recognition rate of 86.54%. Other approaches to emotion intensity are given by Savran et al. [25] and Whitehill et al. [26].

Some of the researchers use a combined dataset to show that their model is not biased to a particular dataset. Ozdemir et al. [27] developed a CNN which was based on the LeNet model. By merging three datasets (JAFFE, KDEF, and custom data) they got training accuracy of 96.43% and validation accuracy of 91.81% in real-time-based facial emotions classification work. Other researcher Ahmed et al. [28] merged eight different datasets and applied augmentation

techniques in their proposed a CNN structure and achieved 96.24% accuracy.

Recently, fusion-based approaches highlighted by the researcher for better accuracy. Kim et al. [29] proposed a hierarchical deep neural network-based FER system in which they fused appearance and geometric-based features. Then Song [30] proposed a feature fusion model based on machine learning and philosophical concepts. Similarly, Park et al. [31] constructed 3D CNN architecture for extracting spatial and temporal features simultaneously. In another research, Chu et al. [32] used multi-layer convolution feature fusion and Zhang et al. [33] proposed mask refined R-CNN that focused on global and detailed information for better results.

Emotions and their intensity play a significant role in various fields. Various approaches of affective computing are already got successful in this task. In this proposed work, we develop an effective approach to find emotions and their intensity. We combined two components CNN especially transfer learning and fuzzy inference system. Although these two components already presented various research in this task. But this kind of fusion of these two models is not done previously. The foundation of this proposed work relies on the specialty of transfer learning and FIS.

Transfer learning has the specialty of feature learning with the assumption that both source and target tasks sufficiently similar. Acquire knowledge from a model and implement it to others gives higher starting accuracy and faster convergence. Moreover, a fuzzy inference system is easy to construct, flexible, and capable to handle vagueness. With the help of fuzzy rules, it maps input values to output. Transfer learning and FIS both are well-established methods and applying in the emotion recognition field for the past couple of years. By grasping its great success in this field, we come up with the idea of an emotion intensity classifier based on transfer learning and FIS to take advantage of their specialty.

Previous emotion intensity researches mainly concentrate on one emotion (happy) also fine-grained categories of intensity are missing [34]. Most intensity-related works are performed on their own dataset [26, 34] or self-annotated datasets [23] by researchers due to the lack of a specific dataset for this task. Also, emotion intensity work based on the FIS model used a large number of face features [17, 18]. To overcome these limitations this work is comprised of two stages: basic emotion classification and subcategory of recognized emotion based on intensity. Pre-trained architecture VGG16 [35] is used for basic emotion classification work and a fuzzy inference system is used to estimating the intensity level of detected emotion. The fusion of these two: pre-trained network-based basic emotion classifier and FIS-based intensity sub classifier work, has not been done previously by other researchers.

The main contributions of this proposed work are as follows:

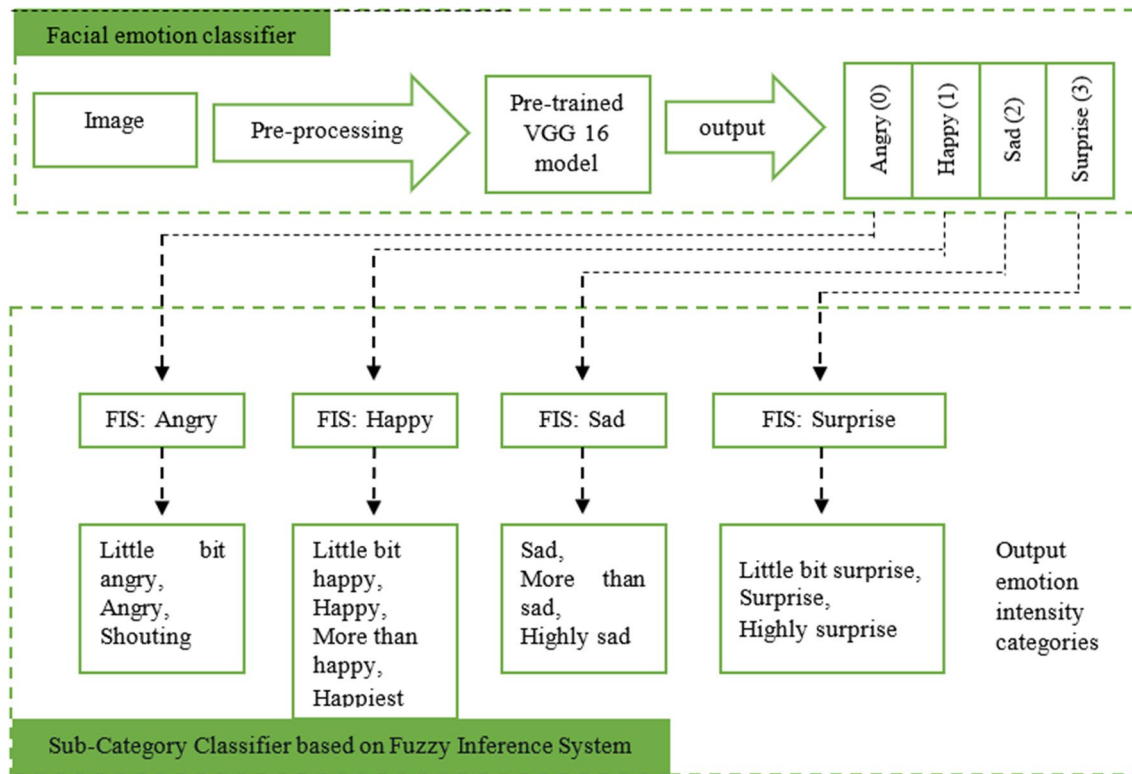


Fig. 1 Flowchart of proposed fusion work: facial emotion intensity classifier

- Aiming at the problem that the emotion intensity work by CNN requires a particular annotated data for this task while FIS-based work requires more face features for accurate prediction. So, the proposed fusion work divided emotion intensity work into two stages: In the first stage it performed basic emotion classification by pre-trained model on the available dataset (CK+, KDEF, and FER 2013) and in the second stage it predicts intensity level of detect emotion by help Fuzzy system with less complexity (i.e. less number of face features) and greater accuracy.
- This work extended emotion intensity work from one emotion happy to more class (happy, sad, surprise, angry). Also, the proposed work is capable of predicting small to peak intensity levels with help of 13 fine-grained categories of intensity.
- We used a combined dataset that contained posed and spontaneous images so variability in data courage real-life implementation of this work and also, we compare the findings of this proposed work to recent related works.

Rest of the paper structured as follows, [2] describes the two proposed modules, section [3] deals with experiment and results, while [4] concentrates on discussion of

experimental results and [5] presents an overview of this work and also highlight its findings.

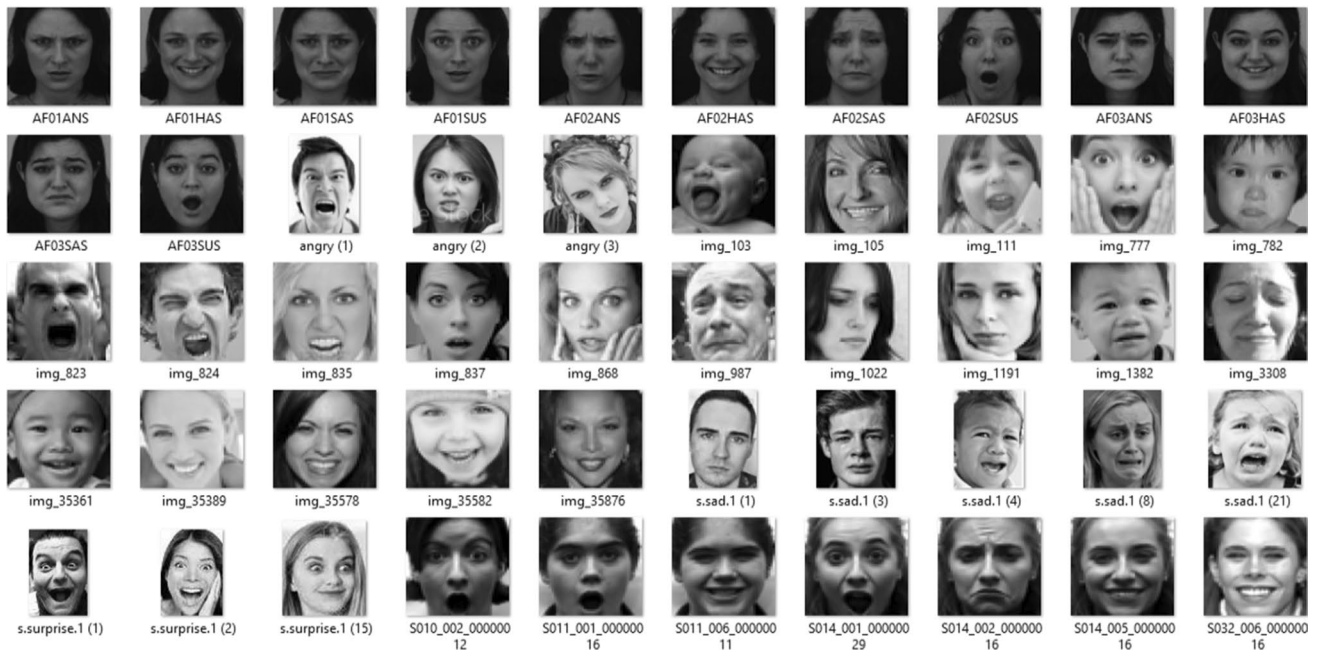
### Proposed Work

This facial emotion intensity classifier work is arranged into two modules: classifier based on pre-trained structure for basic emotions such as happy, sad, angry, surprise, and classifier based on fuzzy inference system for the intensity subcategory of detected emotion. The utility of this model is that it makes emotion subcategory task very easy because after detecting basic emotion by pre-trained structure the subcategory of emotion depends upon selected features of the face like lips, eyes, and in some cases eyebrows. So instead of taking so many feature values or concepts, we can find subcategories of emotion smoothly and precisely. The flowchart of this proposed model is given in Fig. 1.

### Facial Emotion Classifier

#### Collection of Facial Expression Database and Preprocessing

For the first module, we take three databases FER 2013, CK+, and KDEF. FER 2013 dataset [36] contains a total of 35,887 images. Out of 35,887 images, 28,709 for training,



**Fig. 2** Some sample images of dataset

3589 for validation, and 3589 for testing purposes. All images are of size  $48 \times 48$  with 7 emotions. This dataset contains variation in images some images are in a straight position, some of them contain partial faces also there are many images in which the face is cover by hands. This dataset is imbalanced and several images are not correctly annotated.

CK+ dataset introduced by [37] contains 593 image sequences with resolution  $640 \times 490$ . This data set contains posed and spontaneous images of people with ages ranging from 18 to 50 years. CK+ dataset includes 123 different subjects. This dataset also contains the same seven facial expressions as FER 2013 contains.

KDEF (Karolinska Directed Emotional Faces) this dataset contains facial expressions [38] of 35 males and 35 females with 7 major facial expressions (happy, sad, surprise, angry, disgust, afraid, and anger). This dataset contains a total of 4900 images.

Dataset which is used in this paper has been downloaded from Kaggle [39] consists of all three datasets that are mention above with correct annotation. The downloaded database contains 32,900 images of 8 emotions (happy, sad, surprise, angry, disgust, afraid, anger, and neutral). All images are grayscale in PNG format with size  $224 \times 224$ .

In this study, we are focused on 4 emotions (happy, sad, angry, and surprise). From this dataset, we have taken randomly images of 4 emotions for our work. Also, some images are collected from Google. The dataset contains a total of 6937 images. We take 6079 images for training, approximately 1500 images for each class (happy, sad,

angry, and surprise), for validation 436 images and 422 for testing our model. Many researchers used [27, 28] combined datasets for their work. Creating a dataset by collecting images from different sources makes our model effective and unbiased. Figure 2 gives some sample images of this dataset.

**Data preprocessing** The two preprocessing steps for the first module are image resize and image rescaling. Since all images were already in the same size except those which were downloaded from Google. All images were resized to the target size  $224 \times 224$ . Images for training, validation, and testing were loaded using the in-built function ImageDataGenerator provided by Keras API. This function was also used for resizing and rescaling images.

### Basic Emotion Classification by Pre-trained Model

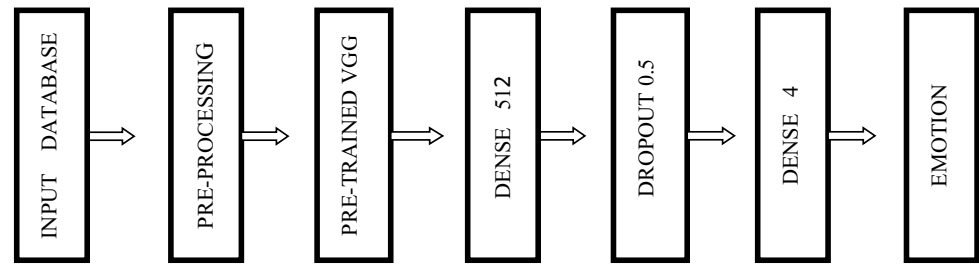
We applied the transfer learning technique using the VGG16 pre-trained model. Transfer learning technique provides weights that are developed for ImageNet image classification tasks. The architecture of the first module is summarized in Fig. 3.

### Sub-category Classifier Based on Fuzzy Inference System

#### Collection of Database and Preprocessing

For second module CK+ and KDEF dataset is used. Since FER 2013 images contain variation in terms of face alignment, face orientation so we did not include this dataset in

**Fig. 3** Architecture of the first module facial emotion classifier



the second module. But to cover a wide range of emotional intensity images were downloaded from Google. After pre-processing on Google images, we built a dataset for this work.

*Preprocessing* CK+ and KDEF dataset contains images that are almost equally orientated and face alignment nearly identical. Also, both datasets comprise only the frontal faces of people. Furthermore, images downloaded from Google required some extra preprocessing effort. We had downloaded frontal face images for four emotions (happy, sad, angry, surprise). Later on, we manually crop images in such a way that images contained face only (similar to images present in CK+ and KDEF) and resize to target size  $224 \times 224$ . Then applied preprocessing steps for the whole dataset which are

- a. Conversion into grayscale—Images were collected from different sources so after resizing them, we need to convert images to grayscale.
- b. Histogram equalization—This is the scheme used for contrast adjustment in the image. Through this method, the intensity of the image is better distributed on the histogram. This equalization method is adequate for both bright and dark images. We applied in-built function “cv2.equalizeHist ()” present in cv2 to reduce data variance.
- c. Face detection and landmark points extraction—Face detection and landmark point’s extraction is a prerequisite step in the FER system. Through face detection, we can find the location of the face in the image. For

this task, we applied a frontal face detector present in the “dlib” library [14, 17]. This face detector is a pre-trained HOG and linear SVM face detector that provides quick and productive results. After detecting face location next, we applied the landmarks predictor which was present in the "dlib" library. It extracted 68 landmark points from the detected face. Figure 4 shows all preprocessing steps.

**Feature Values Estimation/Estimation of Area and Tangent**

Most of the emotion recognition tasks reveal that lips, eyes, and eyebrows are the most informative features. So, various FER related works like Rassadin et al. [14], Farahani et al. [21], Chakraborty et al. [22], and Islam and Loo [16] were based on it. In this work to estimate the lips area and eyes area from the detected face, we considered lips and eyes as elliptical in shape. Also, we calculated lip width and eyebrows tangent of face for emotion intensity-based subcategory work. Lips area and eyes area were calculated by the formula:

Area of Ellipse =  $(3.14 \times \text{length of major axis} \times \text{length of minor axis})$ .

Lip width was calculated by the Euclidean distance (ED) formula and for eyebrows tangent  $(y_2 - y_1)/(x_2 - x_1)$  formula was used. To calculate area and width first of all we calculated normalized Euclidean distance (NED) between two contributing landmark points.



**Fig. 4** Three preprocessing steps: histogram equalization, face detection and landmark points detection

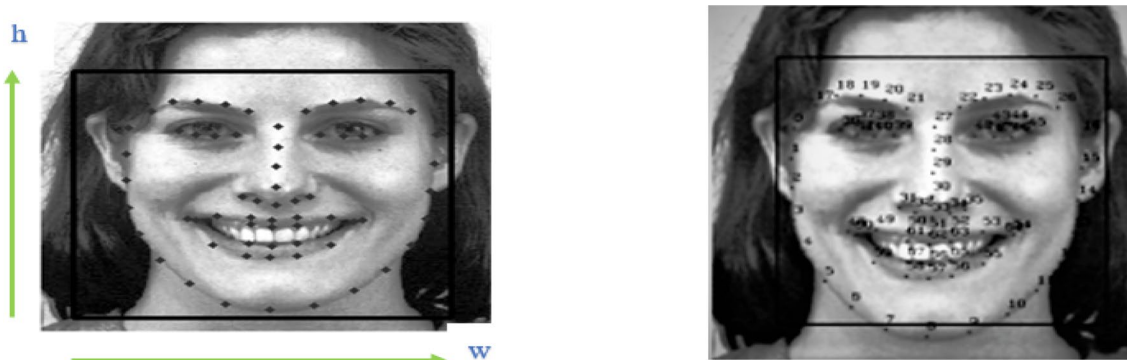


Fig. 5 Height, width of the face and 68 landmark points

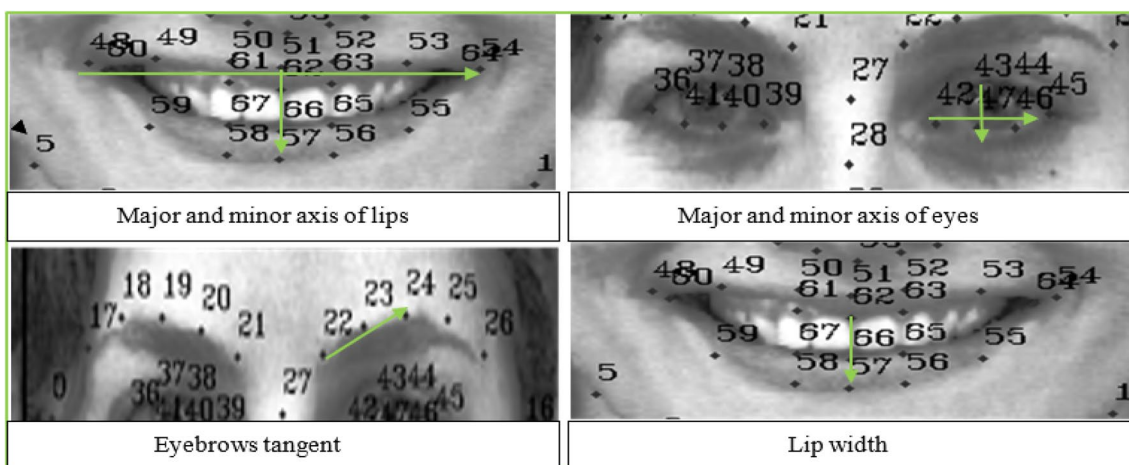


Fig. 6 Face features used in this study

$$NED = (ED/\text{length of face segment}).$$

Euclidean distance cannot use as such because the estimated value varies from image to image depends upon the location of face and area of face segment. To standardize the ED we exerted normalized Euclidean distance as described by Vinola and Vimala Devi [24] in their smile intensity work. Figure 5 shows the height and width of the face and 68 landmark points and Fig. 6 represented all face features used in this study.

**Fuzzy Inference System for Emotion Intensity-Based Sub-categories**

If basic facial emotions are accurately detected by the first module (described in Section “Basic Emotion Classification by Pre-trained Model”) then our second module easily predicts the subcategory of detected emotion. Now to find a subcategory of detected emotion that is based on emotion intensity requires less number of face features. To detect

subcategory of emotion lips, eyes, and in some cases, eyebrows are sufficient as discussed earlier so, lips area, lips width, eyes area, and eyebrow tangent are taken to determine subcategory of emotion based on its intensity.

We constructed four separate fuzzy inference systems for predicting subcategories of four basic emotions. These systems were independent of each other. The reason behind this construction was each emotion subcategory has its interval value with different intensities. For subcategories of each emotion, there were separate fuzzy rules that correspond to the linguistic variable. Range of resultant emotion subcategory values for each emotion taken from 0 to 100.

From the dataset, images corresponding to a class of emotion were taken after that to define the subcategory of emotion we calculated lip area, lip width, eye area, and eyebrow tangent. Table 1 defines the emotions and corresponding face features which were considered under this work. After examining a large dataset with varying emotion intensity, we had drawn a pattern and based on this pattern fuzzy ranges of lips area, lip width, eye area, and eyebrow

**Table 1** Emotions and corresponding face features

Emotion	Face feature	Calculation	Linguistic variables
Angry	Lips	Lip area Lip width	Little bit angry, angry, shouting
	Eyebrow	Eyebrow tangent	
Happy	Lips	Lip area Lip width	Little bit happy, happy, more than happy, happiest
	Eye	Eye area	
Sad	Lips	Lip area Lip width	Sad, more than sad, highly sad
	Eye	Eye area	
Surprise	Lips	Lip area	Little bit surprise, surprise, highly surprise
	Eye	Eye area	

tangent, from lower to higher intensity were defined. The membership function of each fuzzy input is defined by the linguistic variables low, medium, and high. Here triangular and trapezoidal membership functions are used then we defined if .... then type fuzzy rules to predict subcategories for that particular emotion. All fuzzy inference systems were developed in the same manner.

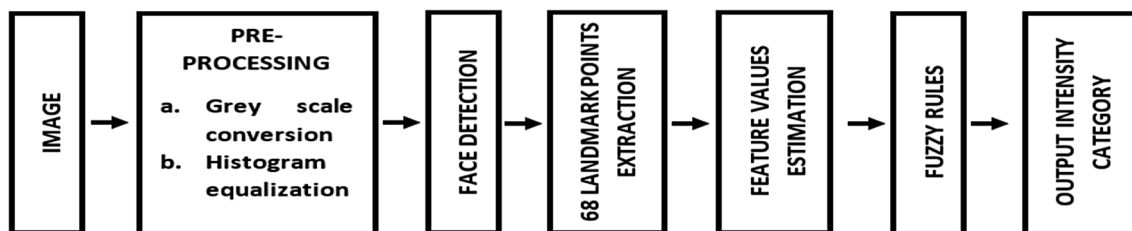
Sample of fuzzy rules for each subcategory of emotion are as follows:

- lip area['low'] & eyebrow tangent['high'], angry['angry'].
- lip area['medium'] & eye area['medium'], surprise['surprise'].
- lip area['low'] & eye area['high'], sad['sad'].
- lip width['low'] & eye area['medium'], happy ['little bit happy'].
- lip area['low'] & eye area['low'], sad ['more than sad'].
- lip area['moderate'] & eye area['moderate'], happy['happy'].
- lip area['low'] & eyebrow tangent['low'], angry ['little bit angry'].
- lip area['high'] & eyebrow tangent['high'], angry['shouting'].

The architecture of the second module is summarized in Fig. 7.

### Experiment and Results

We evaluate the performance of our model through the experiment. The implementation of the proposed work was in Google colab using TensorFlow, Open CV, Dlib, and Scikit Fuzzy libraries. Training, validation, and testing dataset (see “Collection of Facial Expression Database and Preprocessing”) were loaded and preprocessed by the in-built function ImageDataGenerator. Then to train our first module, i.e. basic emotion classifier by pre-trained architecture described in “Basic Emotion Classification by Pre-trained Model” and shown in Fig. 3, we loaded the pre-trained VGG16 model in Keras with taking “include\_top” argument as “false”. To make a prediction we add the first fully connected layer with 512 nodes and “relu” as activation function then dropout layer is introduced in which 50% neurons randomly excluded after this last fully connected layer is added with 4 nodes for classification of 4 basic emotions (happy, sad, angry and surprise) with “softmax” as activation function. “Adam” optimizer with learning rate 0.001 and “categorical cross-entropy” as loss function was picked for this work-frame. Figure 8 shows the model accuracy and model loss. The accuracy of the model for the training dataset was 96.06% and for validation was 81.19% in 100 iterations.



**Fig. 7** Architecture of the second module subcategory classifier based on fuzzy inference system

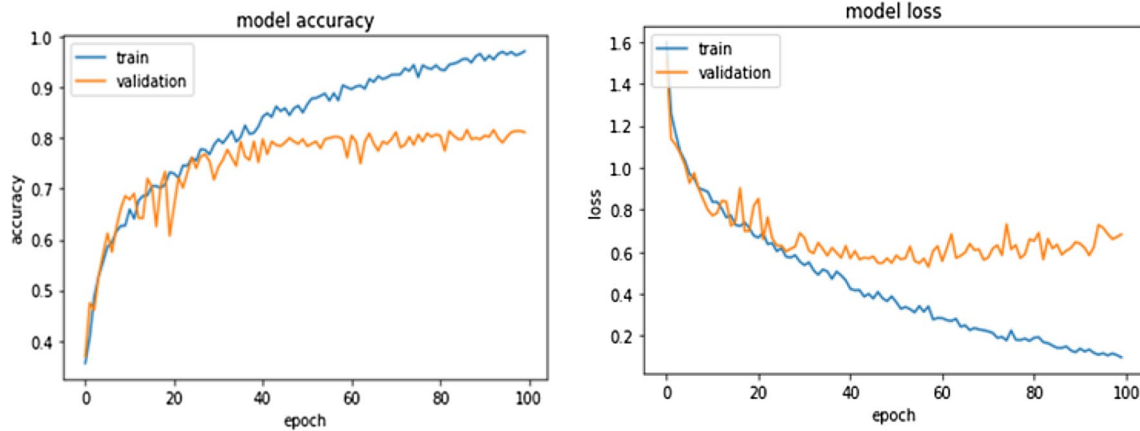


Fig. 8 Model accuracy and loss

	precision	recall	f1-score	support
0	0.82	0.79	0.80	103
1	0.83	0.93	0.88	104
2	0.79	0.70	0.74	105
3	0.89	0.92	0.90	110
accuracy			0.83	422
macro avg	0.83	0.83	0.83	422
weighted avg	0.83	0.83	0.83	422

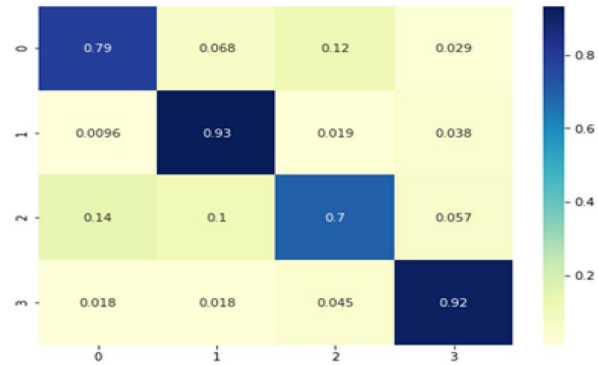


Fig. 9 Precision, recall, f1-score and confusion matrix of classifier

The confusion matrix for the testing data set shows in Fig. 9 and it reveals that the accuracy of the model is 83% for the testing dataset. Precision value shows that if the model predicts a facial emotion is a surprise, it is correct 89% of the time. Also, from the confusion matrix, we conclude that for testing data the recognition rate of both positive emotions is higher than negative emotions. The recognition rate of emotion happy and surprise is 93% and 92%, respectively. While recognition rate for emotion angry and sad is 79% and 70%, respectively.

We picked an image then applied preprocessing steps rescale and resize. After that, first module predicted the basic emotion with the class index value. With the help of this class index value, the system automatically transferred the image to the corresponding fuzzy inference system which was constructed for categorizing that emotion. Before transferring to the FIS system three preprocessing steps were taken histogram equalization, face detection, and landmark point extraction. After examining the location of landmark points in the image, the image was finally transferred to the FIS system. In the FIS system first estimation

process is done (see “[Feature Values Estimation/Estimation of Area and Tangent](#)” and Table 1). Then estimated values pass through the fuzzy system and predicted subcategory of emotion based on fuzzy rules. The pre-trained model employs knowledge and skills from one to another system and the fuzzy system has a grip on uncertainty and impreciseness. Experiment results show that the proposed fusion model: facial emotion intensity classifier first predicts basic emotion class then intensity level-based subclass. Also, it reduces the complexity of this task and increases the performance because if basic emotion is accurately detected by the first module then the intensity of detected emotion depends upon less number of face features (size of mouth opening, eye-opening, and eyebrows tangent). Experiment results for all emotions are shown in Fig. 10.





Fig. 10 Experimental results of all emotions

## Discussion and Future Work

The performance of our model depends upon the prediction accuracy of the basic emotion classifier and preciseness of face detection and landmark point extraction. Here, we used the transfer learning technique, Dlib face detector, and landmark predictor for greater accuracy. To show the performance of the proposed model, images from various datasets (KDEF, CK+, and FER 2013) and also from Google were taken. For the facial emotion recognition task, we just use two preprocessing steps resize and rescale. So, the accuracy of our model fully depends upon the transfer learning technique, fuzzy system and combined dataset which is collected from various sources.

The classification results of the second module emotion intensity classifier will be correct if the first module accurately detects the basic emotion, face detection, and estimation of landmark points. If basic emotions are already detected then the intensity of emotion depends upon the size of mouth opening, eye-opening, and eyebrows tangent as they are prime face features for this task. So, the overall performance of our model depends upon the accuracy of the first module that is 96.06% for training, 81.19% for the validation dataset, and 83% for the testing dataset.

The fuzzy-based second module successfully detects sub-categories of emotion graphically. We perceive that if we examine different images of the same emotion with varying intensity, the model successfully categorizes it. Even if two

images are of the same category then we can do intra-class comparison through their membership values. The proposed system is capable of recognizing small to peak emotion intensity easily and effectively. Through the experiment, we conclude that this proposed work gives significant results for the images taken from different sources. So, overall the proposed fusion model: facial emotion intensity classifier predicts basic emotion class, an intensity value, and also a subcategory of recognized emotion based on its intensity by the graphical way (Fig. 10).

CK+, JAFFE, and KDEF all are posed dataset means data contain no variation in terms of head pose, illumination, and all images have a similar background. So, accuracy reached up to 97% but in the case of spontaneous data (data similar to the real-life situation) like FER2013 researcher got a maximum of 75–76% accuracy. Table 2 summarize the accuracy difference between posed [15, 29, 31, 40–44] and spontaneous datasets [17, 19, 30, 45–47]. Posed datasets always get greater accuracy than spontaneous but are less reliable in real-life applications.

This proposed fusion model: facial emotion intensity classifier predicts combinedly both basic emotion class through module 1 and intensity level-based category through module 2 (Fig. 10). Also, we take a combined dataset in which images are collected from different sources. So, if we compare the basic emotion classifier work (module 1) with other related works (mention in Table 2) having an individual single dataset or a particular number of emotion

**Table 2** Comparison table for facial emotion classifier

Work, year	Dataset	Test accuracy (%)	Model
a. Posed images			
Liew and Yairi (2015) [40]	KDEF	82.40	LBP + Support Vector Machine (SVM)
Li and Lam (2015) [41]	CK +	96.8	Deep Neural Network
Liu et al. (2017) [42]	CK +	97.1	CNN
Alshamsi et al. (2017) [43]	KDEF	90.80	SVM
Zeng et al. (2018) [44]	CK +	95.79	Deep Sparse Autoencoder
Kim et al. (2019) [29]	CK +	96.46	Hierarchical Deep Neural Network
Kondaveeti and Goud (2020) [15]	Combined CK + and JAFFE dataset	83.16	VGG16
Park et al. (2021) [31]	CK +	96.23	3D CNN
b. Spontaneous images			
Kim et al. (2016) [45]	FER2013	73.73	CNN, Network ensemble
Pramerdorfer and Kampel (2016) [46]	FER2013	75.2	CNN, Network ensemble
Georgescu et al. (2019) [47]	FER2013	75.42	CNN + SVM
Bahreini et al. (2019) [17]	Own dataset	83.2	FURIA algorithm
Song (2021) [30]	FER 2013	74	Machine learning and philosophical thinking
Amal et al. (2022) [19]	FER2013	75.1	HOG + CNN
Proposed basic emotion classification work	Combined CK +, KDEF and FER2013 dataset with four emotion classes	83	Pre-trained VGG16 (module 1)

**Table 3** Comparison table for proposed fusion model: facial emotion intensity classifier

Work, year	Emotion classes	Methods
Esau et al. (2007) [18]	Happy, sad, fear, angry	Fuzzy inference system
Whitehill et al. (2009) [26]	Happy	Gabor feature + SVM
Girard et al. (2015) [48]	Smile	Two class SVM, multi class SVM, SVM regression
Dhall et al. (2015) [34]	Happy: neutral, small smile, large smile	Global and local attribute weighted model
Witzig et al. (2019) [23]	Happy: neutral, low, medium, high	CNN + RNN
Vinola and Vimala Devi (2019) [24]	Smile	Fuzzy inference system
Proposed fusion work	<i>Angry</i> : little bit angry, angry, shouting <i>Happy</i> : little bit happy, happy, more than happy, happiest <i>Sad</i> : sad, more than sad, highly sad <i>Surprise</i> : little bit surprise, surprise, highly surprise	Pre-trained VGG16 (module 1) + Fuzzy inference system (module 2)

classes will be unfair. However, proposed basic emotion classification work (module 1) on combined data with four emotion classes got 96.06% training, 81.19% validation, and 83% testing accuracy. Now, we compare the proposed fusion work: facial emotion intensity classifier (Fig. 1) with related previous intensity work in Table 3.

For better communication between human–machine interaction, emotion intensity plays a vital role. In literature, a sufficient amount of work has been done by the researchers in basic emotion classification tasks while on the other hand facial emotion intensity prediction works are limited. The reason behind that no particular labeled dataset is available for this task so, some of the intensity work [26, 34] researchers generate their data set for emotion happy. But for more emotion classes it is time-consuming and expensive to collect all intensity data so, major of previous intensity work [23, 26, 34] concentrates on one emotion happy and also their intensity classes are not sufficient.

To tackle this problem fuzzy-based researches was held in which researchers take a lot of face features [18] (complexity) to define emotion class and intensity level. For instance, Bahreini et al. [17] calculated 54 cosine values for six basic emotion classification work. Vinola and Vimala Devi [24] calculated five Euclidean distances between ten landmark points for smile intensity work only. This proposed fusion-based facial emotion intensity classifier work overcomes all the above-mentioned limitations by generating a classifier with two modules. Due to which no particular additional labeled data set is required for more class emotion intensity work also it reduces the complexity of this task by taking a smaller number of face feature to define the intensity level. The proposed fusion work successfully predicts four emotion classes with 13 emotion subcategories based on intensity. These fine-grained categories of emotion intensity are capable to predict small to peak intensity levels also the graphical output generated by the system is very effective and easy to visualize the outcome. Since this proposed work is also limited to the frontal faces and can be improved by

adding an audio feature also the performance of proposed fusion model depends upon the prediction accuracy of the basic emotion classifier (module 1). So, in the future we will modify our architecture by applying fusion techniques at feature [49], score level [50] and also, we will add more spontaneous images and image preprocessing steps [51] to improve the accuracy. Further, we will apply the data augmentation technique [28], and other pre-trained models [9].

## Conclusion

The main aspiration of this work is to generate an effective way of finding emotions subcategories by fusion of pre-trained network and FIS systems. Our emotion intensity classifier work divides this complicated task into two sub-tasks first basic emotion detection based on CNN especially the transfer learning technique and second subcategory of recognized emotion based on intensity through FIS. One of the important features of this model is that we take a combined dataset that is collected from different sources and our system works effectively on those images that make our model more reliable.

The results of the experiment explore the findings of this work. Experiment results show that the proposed emotion intensity classifier reduces the complexity of this task and enhances the performance by taking the advantage of transfer learning and fuzzy system. The purpose of this proposed work is to find out the improvement opportunity in the emotion intensity work.

**Funding** No funding was received to assist with the preparation of this manuscript.

## Declarations

**Conflict of Interest** The authors have no conflicts of interest to declare that are relevant to the content of this article.

## References

- Chiu I, Pigué O, Diehl-Schmid J, Riedl L, Beck J, Leyhe T, Holsboer-Trachslers E, Kressig RW, Berres M, Monsch AU, Solberger M. Facial emotion recognition performance differentiates between behavioral variant frontotemporal dementia and major depressive disorder. *J Clin Psychiatry*. 2018. <https://doi.org/10.4088/JCP.16M11342>.
- Chu HC, Tsai WWJ, Liao MJ, Chen YM. Facial emotion recognition with transition detection for students with high-functioning autism in adaptive e-learning. *Soft Comput*. 2018;22:2973–99. <https://doi.org/10.1007/S00500-017-2549-Z>.
- Oh KJ, Lee D, Ko B, Choi HJ. A chatbot for psychiatric counseling in mental healthcare service based on emotional dialogue analysis and sentence generation. In: Proc—18th IEEE Int Conf Mob Data Manag MDM 2017. Institute of Electrical and Electronics Engineers Inc.; 2017;371–6. <https://doi.org/10.1109/MDM.2017.64>.
- Wang L, Geng X. Behavioral biometrics for human identification: intelligent applications. *Med Inf Sci Ref*. 2010;505:44–56.
- Saste ST, Jagdale SM. Emotion recognition from speech using MFCC and DWT for security system. In: Proc Int Conf Electron Commun Aeronaut Technol ICECA 2017. Institute of Electrical and Electronics Engineers Inc.; 2017;2017-January: 701–4. <https://doi.org/10.1109/ICECA.2017.8203631>.
- Seli P, Wammes JD, Risko EF, Smilek D. On the relation between motivation and retention in educational contexts: The role of intentional and unintentional mind wandering. *Psychon Bull Rev*. 2016;23:1280–7. <https://doi.org/10.3758/S13423-015-0979-0>.
- Zhao J, Mao X, Chen L. Speech emotion recognition using deep 1D & 2D CNN LSTM networks. *Biomed Signal Process Control*. 2019;47:312–23. <https://doi.org/10.1016/J.BSPC.2018.08.035>.
- Borth D, Ji R, Chen T, Breuel T, Chang SF. Large-scale visual sentiment ontology and detectors using adjective noun pairs. In: MM 2013—Proc 2013 ACM Multimedia Conf. Association for Computing Machinery; 2013; p. 223–32. <https://doi.org/10.1145/2502081.2502282>.
- Sajjanhar A, Wu Z, Wen Q. Deep learning models for facial expression recognition. In: 2018 Int Conf Digit Image Comput Tech Appl DICTA 2018. Institute of Electrical and Electronics Engineers Inc.; 2019; p. 1–6. <https://doi.org/10.1109/DICTA.2018.8615843>.
- Soleymani M, Asghari-Esfeden S, Fu Y, Pantic M. Analysis of EEG signals and facial expressions for continuous emotion detection. *IEEE Trans Affect Comput*. 2016;7:17–28. <https://doi.org/10.1109/TAFFC.2015.2436926>.
- Schindler K, Van Gool L, de Gelder B. Recognizing emotions expressed by body pose: a biologically inspired neural model. *Neural Netw*. 2008;21:1238–46. <https://doi.org/10.1016/J.NEUNET.2008.05.003>.
- Kosti R, Alvarez JM, Recasens A, Lapedriza A. Context based emotion recognition using EMOTIC dataset. *IEEE Trans Pattern Anal Mach Intell IEEE*. 2020;42:2755–66. <https://doi.org/10.1109/TPAMI.2019.2916866>.
- Bhattacharya, S. A Survey on: facial expression recognition using various deep learning techniques. In: Advanced computational paradigms and hybrid intelligent computing. Springer, Singapore. 2022; 1373:619–31. [https://doi.org/10.1007/978-981-16-4369-9\\_59](https://doi.org/10.1007/978-981-16-4369-9_59).
- Rassadin A, Gruzdev A, Savchenko A. Group-Level emotion recognition using transfer learning from face identification. In: ICMI 2017—Proc 19th ACM Int Conf Multimodal Interact. 2017; 2017-Janua:544–8. <https://doi.org/10.1145/3136755.3143007>.
- Kishan Kondaveeti H, Vishal Goud M. Emotion detection using deep facial features. In: Proc IEEE Int Conf Advent Trends Multidiscip Res Innov ICATMRI 2020. 2020; p. 1–8. <https://doi.org/10.1109/ICATMRI51801.2020.9398439>.
- Islam MN, Loo CK. Geometric feature-based facial emotion recognition using two-stage fuzzy reasoning model. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)*. 2014;8835:344–51. [https://doi.org/10.1007/978-3-319-12640-1\\_42](https://doi.org/10.1007/978-3-319-12640-1_42).
- Bahreini K, van der Vegt W, Westera W. A fuzzy logic approach to reliable real-time recognition of facial emotions. *Multimed Tools Appl*. 2019;78:18943–66. <https://doi.org/10.1007/s11042-019-7250-z>.
- Esau N, Wetzel E, Kleinjohann L, Kleinjohann B. Real-time facial expression recognition using a fuzzy emotion model. In: IEEE Int Conf Fuzzy Syst. 2007; p. 1–6. <https://doi.org/10.1109/FUZZY.2007.4295451>.
- Amal VS, Suresh S, Deepa G. Real-time emotion recognition from facial expressions using convolutional neural network with Fer2013 dataset. In: Ubiquitous intelligent systems. Springer, Singapore. 2022; 243:541–51. [https://doi.org/10.1007/978-981-16-3675-2\\_41](https://doi.org/10.1007/978-981-16-3675-2_41).
- Nicolai A, Choi A. Facial emotion recognition using fuzzy systems. In: Proc—2015 IEEE Int Conf Syst Man, Cybern SMC 2015. 2016; p. 2216–21. <https://doi.org/10.1109/SMC.2015.387>.
- Farahani FS, Sheikhan M, Farrokhi A. A fuzzy approach for facial emotion recognition. In: 13th Iran Conf Fuzzy Syst IFSC 2013. 2013; p. 1–4. <https://doi.org/10.1109/IFSC.2013.6675597>.
- Chakraborty A, Konar A, Chakraborty UK, Chatterjee A. Emotion recognition from facial expressions and its control using fuzzy logic. *IEEE Trans Syst Man Cybern Part A Syst Hum*. 2009;39:726–43. <https://doi.org/10.1109/TSMCA.2009.2014645>.
- Witzig P, Kennedy J, Segalin C. Smile intensity detection in multiparty interaction using deep learning. In: 2019 8th Int Conf Affect Comput Intell Interact Work Demos, ACIIW 2019. 2019; p. 168–74. <https://doi.org/10.1109/ACIIW.2019.8925261>.
- Vinola C, Vimala DK. Smile intensity recognition in real time videos: fuzzy system approach. *Multimed Tools Appl*. 2019;78:15033–52. <https://doi.org/10.1007/s11042-018-6890-8>.
- Savran A, Sankur B, Taha BM. Regression-based intensity estimation of facial action units. *Image Vis Comput*. 2012;30:774–84. <https://doi.org/10.1016/J.IMAVIS.2011.11.008>.
- Whitehill J, Littlewort G, Fasel I, Bartlett M, Movellan J. Toward practical smile detection. *IEEE Trans Pattern Anal Mach Intell*. 2009;31:2106–11. <https://doi.org/10.1109/TPAMI.2009.42>.
- Ozdemir MA, Elagoz B, Alaybeyoglu A, Sadighzadeh R, Akan A. Real time emotion recognition from facial expressions using CNN architecture. In: TIPTEKNO 2019—Tip Teknol Kongresi. 2019; p. 529–32. <https://doi.org/10.1109/TIPTEKNO.2019.8895215>.
- Ahmed TU, Hossain S, Hossain MS, Ul Islam R, Andersson K. Facial expression recognition using convolutional neural network with data augmentation. In: 2019 Jt 8th Int Conf Informatics, Electron Vision, ICIEV 2019 3rd Int Conf Imaging, Vis Pattern Recognition, icIVPR 2019 with Int Conf Act Behav Comput ABC 2019. 2019; p. 336–41. <https://doi.org/10.1109/ICIEV.2019.8858529>.
- Kim JH, Kim BG, Roy PP, Jeong DM. Efficient facial expression recognition algorithm based on hierarchical deep neural network structure. *IEEE Access*. 2019;7:41273–85.

30. Song Z. Facial expression emotion recognition model integrating philosophy and machine learning theory. *Front Psychol*. 2021;12:759485. <https://doi.org/10.3389/fpsyg.2021.759485>.
31. Park SJ, Kim BG, Chilamkurti N. A robust facial expression recognition algorithm based on multi-rate feature fusion scheme. *Sensors*. 2021;21(21):1–26.
32. Chu J, Guo Z, Leng L. Object detection based on multi-layer convolution feature fusion and online hard example mining. *IEEE Access*. 2018;6:19959–67.
33. Zhang Y, Chu J, Leng L, Miao J. Mask-refined R-CNN: a network for refining object details in instance segmentation. *Sensors (Switzerland)*. 2020;20(4):1010.
34. Dhall A, Goecke R, Gedeon T. Automatic group happiness intensity analysis. *IEEE Trans Affect Comput IEEE*. 2015;6:13–26. <https://doi.org/10.1109/TAFFC.2015.2397456>.
35. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv*: 2014:1409–1556.
36. Goodfellow IJ, Erhan D, Carrier PL, Courville A, Mirza M, Hamner B, Cukierski W, Tang Y, Thaler D, Lee DH, Zhou Y, Ramezani C, Feng F, Li R, Wang X, Athanasakis D, Shave-Taylor J, Mikolajczyk M, Park J, Ionescu R, Popescu M, Grozea C, Bergstra J, Xie J, Romaszko L, Xu B, Chuang Z, Bengio Y. Challenges in representation learning: a report on three machine learning contests. In: *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)*. 2013;8228 LNCS:117–24. [https://doi.org/10.1007/978-3-642-42051-1\\_16](https://doi.org/10.1007/978-3-642-42051-1_16).
37. Lucey P, Cohn JF, Kanade T, Saragih J, Ambadar Z, Matthews I. The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression. In: *2010 IEEE Comput Soc Conf Comput Vis Pattern Recognit—Work CVPRW 2010*. 2010; p. 94–101. <https://doi.org/10.1109/CVPRW.2010.5543262>.
38. Lundqvist D, Flykt A, Öhman A. The Karolinska directed emotional faces—KDEF. Department of Clinical Neuroscience, Psychology Section, Karolinska Institutet; 1998.
39. Vaidya S. Corrective re-annotation of FER-CK+-KDEF 2020. Available at: <https://www.kaggle.com/sudarshanvaidya/corrective-reannotation-of-fer-ck-kdef>. Accessed 3 May 2021.
40. Liew CF, Yairi T. Facial expression recognition and analysis: a comparison study of feature descriptors. *IPSPJ Trans Comput Vis Appl*. 2015;7:104–20. <https://doi.org/10.2197/ipsjtcva.7.104>.
41. Li J, Lam EY. Facial expression recognition using deep neural networks. In: *IST 2015—2015 IEEE Int Conf Imaging Syst Tech Proc*. IEEE; 2015; p. 1–6. <https://doi.org/10.1109/IST.2015.7294547>.
42. Liu X, Kumar BVKV, You J, Jia P. Adaptive deep metric learning for identity-aware facial expression recognition. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2017; p. 522–31. <https://doi.org/10.1109/CVPRW.2017.79>.
43. Alshamsi H, Kepuska V, Meng H. Real time automated facial expression recognition app development on smart phones. In: *2017 8th IEEE Annu Inf Technol Electron Mob Commun Conf IEMCON 2017*. 2017; p. 384–92. <https://doi.org/10.1109/IEMCON.2017.8117150>.
44. Zeng N, Zhang H, Song B, Liu W, Li Y, Dobaie AM. Facial expression recognition via learning deep sparse autoencoders. *Neurocomputing [Internet]*. 2018;273:643–9. <https://doi.org/10.1016/j.neucom.2017.08.043>.
45. Kim BK, Dong SY, Roh J, Kim G, Lee SY. Fusing aligned and non-aligned face information for automatic affect recognition in the wild: a deep learning approach. In: *IEEE Comput Soc Conf Comput Vis Pattern Recognit Work*. 2016; p. 1499–508. <https://doi.org/10.1109/CVPRW.2016.187>.
46. Pramerdorfer C, Kampel M. Facial Expression Recognition using Convolutional Neural Networks: State of the Art. 2016; *arXiv*: <http://arxiv.org/abs/1612.02903>.
47. Georgescu MI, Ionescu RT, Popescu M. Local learning with deep and handcrafted features for facial expression recognition. *IEEE Access IEEE*. 2019;7:64827–36. <https://doi.org/10.1109/ACCESS.2019.2917266>.
48. Girard JM, Cohn JF, De La Torre F. Estimating smile intensity: A better way. *Pattern Recognit Lett*. 2015;66:13–21. <https://doi.org/10.1016/j.patrec.2014.10.004>.
49. Leng L, Li M, Kim C, Bi X. Dual-source discrimination power analysis for multi-instance contactless palmprint recognition. *Multimed Tools Appl [Internet]*. 2017;76(1):333–54. <https://doi.org/10.1007/s11042-015-3058-7>.
50. Leng L, Zhang J. PalmHash code vs. palmPhasor code. *Neurocomputing [Internet]*. 2013;108:1–12. <https://doi.org/10.1016/j.neucom.2012.08.028>.
51. Pitaloka DA, Wulandari A, Basaruddin T, Liliana DY. Enhancing CNN with preprocessing stage in automatic emotion recognition. *Proc Comput Sci*. 2017;116:523–9. <https://doi.org/10.1016/j.procs.2017.10.038>.

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