Conventional Feature Engineering and Deep Learning Approaches to Facial Expression Recognition: A Brief Overview

Shubh Lakshmi Agrwal[1,2]([✉]), Sudheer Kumar Sharma[2], and Vibhor Kant[3]

[1] Manipal University Jaipur, Jaipur, India
shubhu.agarwal@gmail.com
[2] LNM Institute of Information Technology, Jaipur, India
[3] RGSC, Banaras Hindu University, Varanasi, India

Abstract. Facial expression recognition (FER) is vital in pattern recognition, artificial intelligence, and computer vision. It has diverse applications, including operator fatigue detection, automated tutoring systems, music for mood, mental state identification, and security. Image data collection, feature engineering, and classification are vital stages of FER. A comprehensive critical review of benchmarking datasets and feature engineering techniques used for FER is presented in this paper. Further, this paper critically analyzes the various conventional learning and deep learning methods for FER. It provides a baseline to other researchers about future aspects with the pros and cons of techniques developed so far.

Keywords: Facial expression recognition · Feature engineering · Conventional learning · Deep learning · Face expression dataset

1 Introduction

Facial expressions are crucial for social communication. Verbal and nonverbal communication are standard. Facial expressions communicate non-verbally. Mehrabian [40] revealed that 55% of information passes between people through facial expressions, 38% via voice, and 7%via language [66]. Facial expression recognition has evolved into an outstanding and demanding field of computer vision. Disgust, anger, happiness, fear, surprise, and sadness are fundamental emotions [13]. Humans are highly skilled at identifying a person's emotional state; a computer would have difficulty doing so. It is caused by a variation in occlusion, head postures, changes in lighting, and computing complexity. FER applications include operator tiredness detection, [77], automobile, healthcare, automated tutoring systems [67], mental state recognition [39], security [6], music for mood [12], and rating products or services in banks, malls, and showrooms. With the help of a FER, users can also study how well students interact in a classroom or talk with teachers. [56]. FER inbuilt mobile applications can help visually impaired persons (VIPs) to communicate daily. FER systems can detect the driver's fatigue

state and stress level to make better decisions about driving safely. Facial image acquisition, pre-processing, feature engineering, training, and classification are typical FER stages. The Fig. 1 depicts face expression recognition steps. Pre-processing is used to remove noise. Feature engineering extracts distinct visual characteristics. The popular feature engineering techniques are, Histogram of Gradient(HOG) [10], Local Directional Pattern (LDP) [23], Gabor filters [61], Local Binary Patterns (LBP) [52], Principal Component Analysis (PCA) [2], Independent Component Analysis (ICA), and Linear Discriminant Analysis(LDA) [5]. Extracted features are utilized for training a classifier using expression class labels. FER approaches are deep learning and conventional learning based on feature engineering. In deep learning huge number of examples and images are used to learn and tune feature extraction parameters, while conventional learning uses algorithms to extract hand-crafted features. Deep learning classifiers contain a sigmoid or softmax layer on the classification stage with Fully connected layers. K-nearest neighbor (KNN)and support vector machine(SVM)are well-known classifiers in conventional learning. The FER system's accuracy depends on captured data variability, feature extraction, classification, and fine-tuning. Model inference time depends on camera resolution, feature engineering, classifier, and hardware computation capabilities.
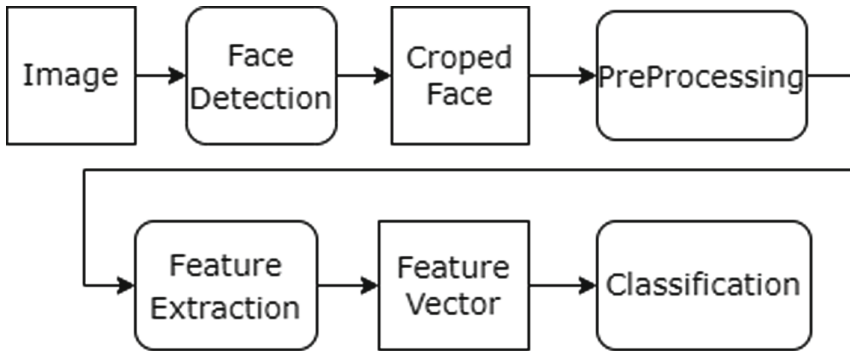

Fig. 1. Different steps of facial expression recognition system

This work primarily concerns various FER approaches, with three primary processes: pre-processing, feature engineering, and classification. This paper also demonstrates the benefits of different FER methods and a performance analysis of various FER methods. Only image-based FER approaches are used in this work for the literature review; video-based FER techniques are not used. FER systems often deal with illumination fluctuations, skin tone variations, lighting variations, occlusion, and position variations. This work also provides a vital research suggestion for future FER research. The remaining research paper is organized into five 6 sections, including an introduction. Section 2 represents the related research work, including state-of-the-art for FER. Section 3 lists the

most often used benchmarking datasets for FER. Section 4 provides an overview of FER feature engineering. Section 5 compares the performance of different FER systems. Finally, Sect. 6 offers a conclusion.

2   Related Work

FER has a wide range of applications in computer vision. Because of differences in position, illumination, scale, and orientation, recognizing facial expressions can be difficult. The primary goal of feature engineering is to find robust features that can improve the robustness of expression recognition. The feature extraction and classification stages are critical in FER. There are two kinds of feature extraction: geometric and appearance-based. Geometrically-based feature extraction includes the eye, mouth, nose, brow, ear, and other facial components, whereas appearance-based feature extraction includes the exact region of the face [66].

Abdullah et al. [1] reduced the face picture into a small feature set called eigenface and utilized PCA to extract facial features like eigenfaces into a class of finite feature descriptions. Yadav et al. [70] extracted facial features using Gabor filters and two-dimensional PCA. ICA is used to identify characteristics from statistically independent local faces [59]. Lee et al. [30] used ICA to extract statistically autonomous features from local face parts in various facial expressions. Mehta and Jadhav [41] classified human emotions using the Gabor filter. Islam et al. [22] used HOG and LBP to extract local characteristics. LBP features are easy to compute. ICA is less tolerant of illumination fluctuations than LBP. Edge pixels are needed to extract face features from an image. Local Directional Pattern (LDP) shows visual gradients. In FER, LDP represents gradient-based properties of the local face in the pixel's eight prime directions [23].

In classic LDP features, the highest edge strengths determine binary values, which vary by experiment. LDP ignores a pixel's direction strength sign, differentiating edge pixels with comparable strengths but opposite signs. Uddin et al. [60] overcame this LDP problem by grouping pixels' major edge strengths in decreasing order and using their signs to build stable features. Many recent attempts have been made to recognize facial expressions from videos or images using deep learning. To learn appearance features from video frames and geometric features from raw face landmarks, Jung et al. [24] merged two deep learning-based models. Then, a joint learning method was used to connect the two models' outputs. Zeng et al. [75] improved the performance by incorporating hand-picked features into the deep network training. Recently several deep learning methods have been developed for FER and applied in real-time images. Wang et al. [63] introduced Region Attention Network (RAN) for pose variant and occluded face FER. In this paper, region-biased loss and region attention mechanisms are employed to capture the importance of pose variant and occluded facial images. Wang et al. [62] proposed a ResNet-18 CNN model in which uncertainties caused by low-quality images are suppressed by CNN architecture Self-Cure Network (SCN). Li et al. [32] proposed a model that includes an attention mechanism in CNN to recognize expression from a partially occluded face.

3   Review Analysis of Facial Expression Dataset

This section describes FER benchmark datasets. The summary of these datasets, i.e., collection condition, environment challenges, expression distribution, and the number of images and subjects, is shown in Table 1. In the CK+ dataset, training, testing, and validation sets are not specified. Due to non-uniform expressive representation, MMI contains substantial interpersonal discrepancy. The JAFFE dataset has fewer samples per subject expression. AFEW is a multi-model, temporal dataset containing environmental conditions. CMU Multi-PIE and BU-3DFE examine multi-view face expressions.

Table 1. Benchmarking datasets for facial expression recognition

Dataset	Images, subjects	Expression	Environment challenges	Collection
JAFFE [25]	213, 10	6 basic exp + neutral	Pos	Lab
CK+ [37]	593, 123	6 basic exp, Natural, contempt	Sp, Pos	Lab
SFEW [11]	1766, 95	6 basic exp	Sp, Pos	Movie
MMI [48]	740, 25	6 basic exp	Pos	Lab
CMU-MultiPIE [17]	755370, 337	Neutral, disgust, smile surprised, scream, squint	Pos	Lab
FER 2013 [8]	35887	6 basic exp	Sp, Pos	Web
RAF-DB [31]	29672	6 basic, 12 compound exp	Sp, Pos	Internet
EmotioNet [14]	1,000,000	6 basic, 17 compound exp	Sp, Pos	Internet
AffectNet [43]	450, 000	6 basic exp	Sp, Pos	Internet
ExpW [78]	91, 793	6 basic exp	Sp, Pos	Internet
RaFD [29]	1608,67	6 basic, 12 compound exp	Pos	Lab
BU3DFE [73]	2500, 100	6 basic exp + neutral	Pos	Lab
TFD [55]	112234	6 basic exp + neutral	Pos	Lab

4   Review of Feature Engineering Technique

FER accuracy depends on feature engineering. Feature engineering can be hand-picked or deep-learned. Single-task learning (STL) includes hand-picked features, whereas deep learning techniques are iterative. FER's traditional feature engineering methodologies are as follows:

4.1   Gaussian Mixture Model

Gaussian Mixture Model groups data into a cluster that is distinct from each other. A distribution models data points within a cluster. A weighted total of Gaussian functions can approximate many probability distributions. A Gaussian

mixture model is the sum of k component Gaussian densities for vector x, as shown in Eq. 1.

$$p(x) = \sum_{j=1}^{k} w_j p(x|j) \tag{1}$$

where x is a data vector of D dimension; $w_j = 1, 2 \ldots k$, are the weights of the mixture; $p(x|j) =$ Gaussian Density Model for j^{th} Component,

Gaussian one-dimensional probability density function is represented in Eq. 2.

$$G(X|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \tag{2}$$

Here μ represents the mean, and σ^2 represents the distribution variance.

Multivariate Gaussian distribution probability density function is given by Eq. 3 [19].

$$G(X|\mu, \Sigma) = \frac{1}{\sqrt{2\pi^d |\Sigma|}} exp(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(X-\mu)) \tag{3}$$

where μ is a d dimensional vector denoting the mean of the distribution and Σ is the $d \times d$ covariance matrix. The Expectation-Maximization (EM) method estimates model parameters.

4.2   Local Binary Pattern (LBP) Based Features

LBP captures local spatial patterns and the contrast in the facial image. LBP labels image pixels by thresholding the nearby pixel and gives a binary number [47]. LBP is computed in four steps as follows:

– For each pixel (x, y) in an image I, P neighboring pixels are chosen at a radius R.
– Intensity difference of the P adjacent pixels is determined.
– Positive intensity differences are assigned one (1) and negative intensity differences are assigned zero (0).
– Convert the P-bit vector to decimal. LBP descriptor is shown in Eq. 4.

LBP operator $LBP_{P,R}$, here subscript represents the operator used in (P, R) neighborhood.

$$LBP(P, R) = \sum_{p=0}^{p-1} f(i_p - i_c) 2^p \tag{4}$$

where P denotes the number of neighboring pixels chosen at a radius R. i_c and i_p represent the intensity of the center and neighboring pixel, respectively. Thresholding function f is as follows:

$$f(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases} \tag{5}$$

The LBP histogram is defined as:

$$H_j = \sum_{x,y} I\{f_I(x,y) = j, \qquad j = 0, \ldots, n-1 \tag{6}$$

where n is the number of labels created by the LBP operator.

$$I(M) = \begin{cases} 1, & \text{if M is true} \\ 0, & \text{if M is false} \end{cases} \tag{7}$$

Different-sized image patches are normalized using Eq. 8.

$$N_j = \frac{H_j}{\sum_{k=0}^{n-1} H_k} \tag{8}$$

4.3   Gabor Filter Feature Extraction Technique

Edges and texture are essential features in the face image. The convolution of the face image and Gabor filter kernel creates these features. Gabor is an illumination-invariant Gaussian sinusoidal. Gabor filter kernel [65] is defined in Eq. 11. The Gabor filter components are the following: ϕ (Phase), λ (Wavelength), θ (Orientation) specify the number of cycles, angle of the normal to the sinusoidal plane, and offset of a sinusoidal. The frequency bandwidth of Gabor is:

$$b = \log_2 \frac{(\sigma/\lambda)\pi + \sqrt{\log 2/2}}{(\sigma/\lambda)\pi - \sqrt{\log 2/2}} \tag{9}$$

$$\frac{\sigma}{\lambda} = (1/\pi)\sqrt{\log 2/2}\frac{2^b + 1}{2^b - 1} \tag{10}$$

The bandwidth b affects σ value. Convolution of Face image $I(x,y)$ with Gabor kernel $\Psi(\theta, \lambda, \gamma, \phi)$ produces Gabor texture-edge features, shown in Eq. 13 [18]. Gabor kernel $\Psi(\theta, \lambda, \gamma, \phi)$ is composite number as shown in Eq. 14. Gabor real (GI_R) and imaginary (GI_{Im}) components are created by convolution between Gabor kernel Ψ and image $I(x,y)$ for real $R(\Psi)$ and imaginary $\text{Im}(\Psi)$ as shown in Eq. 15 and 16. Equation 17 shows amplitude features $G(x,y)$ of the Gabor kernel. Gabor filter has the problem of redundant features and high dimensions; PCA and ICA can fix this issue.

$$\Psi_{\theta,\lambda,\gamma,\phi}(x,y) = \exp\left(-\frac{a'^2 + \gamma^2 b'^2}{2\sigma^2}\right)e^{j\frac{2\pi a'}{\lambda}} \tag{11}$$

Here, a', b' are direction coefficients and θ represents projection angle.

$$a' = a\cos\theta + b\sin\theta \quad \text{and} \quad b' = a\cos\theta + b\sin\theta \tag{12}$$

$$GI = I(x,y) * \Psi(\theta, \lambda, \gamma, \phi) \tag{13}$$

$$\Psi(\theta, \lambda, \gamma, \phi) = R(\Psi(\theta, \lambda, \gamma, \phi) + Im(\Psi(\theta, \lambda, \gamma, \phi)) \tag{14}$$

$$GI_R(\theta, \lambda, \gamma, \phi) = I(x, y)) * R(\Psi(\theta, \lambda, \gamma, \phi)) \tag{15}$$

$$GI_{Im}(\theta, \lambda, \gamma, \phi) = I(x, y) * Im(\Psi(\theta, \lambda, \gamma, \phi)) \tag{16}$$

$$GF(\theta, \lambda) = (GI_R(\theta, \lambda, \gamma, \phi)^2 + (GI_{Im}(\theta, \lambda, \gamma, \phi)^2)^{1/2} \tag{17}$$

4.4 SIFT-Scale Invariant Feature Transform

SIFT Features are invariant to the scale of the image. The steps for calculating SIFT features are following.

1. Scale-space extrema detection: Gaussian difference finds scale- and rotation-invariant nearest points. Scale and image location are computed.
2. Key point localization: Only solid and fascinating points are selected based on intensity.
3. Orientation assignment: Key points are formed based on the gradient's direction.
4. Key point descriptor: SIFT descriptions around important points are used to describe the local appearance of key points.
5. Keypoint matching: Two images' nearest neighbors are matched.

4.5 Histogram of Oriented Gradient (HOG) Feature Extraction

Facial characteristics vary. A woman's face is rounder than a man's, which helps distinguish gender. HOG extracts picture curvature direction. Edge directions define the shape and local appearance [10]. The image is divided into blocks, and HOG features are computed for each block. All HOG features are integrated into one vector. HOG computation process: Calculate image gradient. For a face image F,

$$F_x = F(r, c+1) - F(r, c-1), \qquad F_y = F(r-1, c) - F(r+1, c) \tag{18}$$

here r and c represent rows and columns, respectively.

The magnitude (G) and orientation (θ) of the gradient is computed by

$$\mid G \mid = \sqrt{F_x{}^2 + F_y{}^2} \quad \text{and} \quad \theta = \tan^{-1}\frac{F_y}{F_x} \tag{19}$$

Orientation range (0–$360°$) for signed gradients and (0–$180°$) for unsigned gradients. After determining the size and orientation of each cell's pixel, the histogram is normalized using a block pattern. Combining HOG features creates a feature vector.

4.6   Discrete Wavelet Transform (DWT)

DWT can be computed by first evaluating 1D-DWT on rows of the 2D image matrix, then on the column of evaluated 1D-DWT. LL(low frequency), LH, HL, and HH (high frequency) represent approximation, horizontal, vertical, and diagonal frequency blocks, respectively, in DWT. LL block approximates low-resolution images by deleting extraneous details. Low-frequency band (LL) smooths input image while high frequency creates edge patterns [54]. Iteratively utilizing 2D-DWT on the LL band helps to reduce feature size.

4.7   Principle Component Analysis (PCA)

PCA finds correlations across attributes and uses a strong variance pattern to reduce data dimensions. In PCA, the given image is subtracted from the mean; the covariance matrix is calculated using FM^T, then eigenvalues and eigenvectors are calculated. Eigenvectors that match up with certain high-magnitude eigenvalues at a certain significant level are essential information about the image's variance. Equation 20 is used to figure out the PCA significance level.

$$\varepsilon = \frac{\sum_{i=1}^{m} \lambda_i}{\sum_{i=1}^{n} \lambda_i} \qquad m \leq n \quad and \quad 0 \leq \varepsilon \geq 1 \tag{20}$$

here λ_i depicts the eigen value of ith order in order of amplitude and $m \leq n$.

4.8   Deep-Learning Feature Engineering

Recent research has emphasized deep learning. Deep learning features are extracted using a convolution neural network (CNN). A DNN was proposed to retrieve patterns from high-dimensional data [27]. DNNs train slowly and overfit. Deep Belief Network [35]is used to tackle DNN challenges, with Restricted Boltzmann Machine (RBM) for training features [42]. A joint learning algorithm is used to combine geometry and appearance features [24].

5   Performance Analysis of Different FER Systems

The performance analysis of this review is based on the pre-processing, recognition accuracy on various datasets, feature extraction methods, contribution, and advantages of different FER techniques. Table 2 shows a comparative analysis of facial expression recognition techniques to better understand how complicated and accurate the method is.

5.1   Conventional Learning-Based FER Analysis

LBP feature extraction and paired classification outperformed JAFFE with 99.05 accuracy [9]. Pairwise classifiers select features by class pair. Feature extraction is more dependable because it doesn't rely on manually or automatically assigned

Table 2. Performance comparison based on different hand-picked feature engineering (conventional learning) and deep learning approaches for facial expression recognition.

	Authors	Feature extraction, classification	Dataset	Accuracy (%)
Conventional learning	Islam et al. [22]	HOG+LBP, ANN	JAFFE CK+	93.51 99.67
	Kumar et al. [28]	WPLBP, SVM	JAFFE	98.15
	Cossetin et al. [9]	LBP, Pairwise Classifier	JAFFE	99.05
	Mehta and Jadhav [41]	Log Gabor-+PCA-CNN	JAFFE	93.57
	Mollahosseini et al. [42]	Intra Face DNN	FER-2013	66.4
	Xu and Zhao [69]	LBP-CNN	FER-2013	94.73
	Aghamaleki et al. [3]	LBP+Sobel-multistream CNN	CK+	98.18
	Shan et al. [53]	Haar like features, CNN	JAFFE	76.74
	Nazir et al. [44]	HOG+DCT, KNN	MMI	99.4
	Kar et al. [26]	HOG+PCA+LDA	CK+	99.2
	Liu et al. [36]	LBP+HOG	CK+	99.6
	Tsai and Chang [57]	DCT+Gabor Filter, SVM	JAFFE	97.10
	Ryu et al. [50]	LDTP, SVM	JAFFE MMI	94.8 99.8
	Nigam et al. [46]	DWT+HOG, SVM	JAFFE	71.43
Deep learning	Alam et al. [4]	Sparse deep recurrent model (S-DSRN)	CK+	99.23
	Liu et al. [35]	Boosted deep belief network (BDBN)	CK+ JAFFE	96.70 93
	Liu et al. [33]	3D CNN with deformable action parts (3D-CNN-DAP)	CK+	92.40
	Liu et al. [34]	CNN +SVM	CK+	93.50
	Zhang et al. [76]	Deep spatial-temporal networks	CK+	98.5
	Kim et al. [27]	Hierarchical DNN	JAFFE CK+	91.27 96.46
	Happy et al. [21]	Salient Facial Patches	CK+	94.09
	Xie and Hu [68]	DCMA-CNN	CK+	93.46
	Yang et al. [72]	De-expression Residue Learning (DeRL)	CK+	97.3
	Yu et al. [74]	Deeper Cascaded Peak- piloted Network (DCPN)	CK+	99.6
	Pramerdorfer et al. [49]	CNN	FER-2013	75.2
	Wang et al. [64]	Pose generative-adversarial network (PGAN)	FER-2013 JAFFE	71.1 95.7
	Mahmoudi et al. [38]	CNN based bilinear model	FER 2013	77.81
	Saurav et al. [51]	Dual integrated convolution neural network (DICNN)	FER 2013 CK+	72.77 96.68
	Liu et al. [35]	Boosted DeepBelief Net (BDBN)	JAFFE CK+	93 96.7
	Turan et al. [58]	Soft Locality Preserving Map (SLPM)	CK+ JAFFE	96.10 91.8
	Hamester et al. [20]	Multi-channel Convolutional Neural Network (MCCNN)	JAFFE	95.8
	Gonzalez [16]	CNN+SVM linear	JAFFE	92
	Ng et al. [45]	CNN	SFEW	55.6
	Cai et al. [7]	CNN	SFEW	59.41
	Zhao et al. [79]	Global multi-scale and local attention network (MA-Net)	SFEW	59.4
	Fan et al. [15]	Fusion of RNN and 3D CNN	AFEW	59.02
	Yan et al. [71]	CNN and bidirectional RNN	AFEW	56.66

fiducial points. Islam et al. [22] used HOG, LBP features, artificial neural network(ANN) classifier to get 99.67% accuracy on CK+ dataset. Feature fusion gives promising results, and ANN employs the limited memory (L-BFGS) technique for weight optimization but the dimension increases. Principle Component Analysis(PCA) is used to reduce dimension. Ryu et al. [50] extracted features via Local Directional Ternary Pattern(LDTP), used classifier as Support vector machine (SVM) and got accuracy 99.8% on MMI dataset.

5.2   Deep Learning-Based FER Analysis

Deeper Cascaded Peak-piloted Network (DCPN) [74] is achieved best accuracy 99.6% on the CK+ dataset, which is greater than other approaches in Table 2. Mahmoudi et al. [38] developed a CNN-based bilinear model and outperformed on FER-2013 dataset (unconstrained dataset) with 77.81% accuracy.

6   Conclusions

This paper present review of different features engineering techniques and provides detailed analysis including pros and cons of each techniques and comparative study of benchmarking dtaset. The techniques are categorized into conventional learning and deep-learning. Conventional learning including LBP, PCA, Gabor filter, HOG, DCT, DWT etc. feature extraction techniques while deep learning includes convolution neural networks and its variant for facial expression recognition. FER systems based on conventional learning and deep learning are presented with the help of bench-marking datasets based accuracy. Hybrid features provide a better recognition rate as compared to single features. This paper analyzed the different FER techniques according to pre-processing, feature engineering, classification, recognition accuracy, and critical contributions. The success of the FER approach depends on pre-processing of the facial images due to illumination and prominent feature engineering. Deep learning model performance is significantly better than conventional learning for real-time datasets but needs a huge amount of datasets and variability of images. The performance of these algorithms improves with the amount of the dataset. JAFFE and CK+ datasets are most frequently used in FER systems, but they do not contain all variability of real-time images. Although much research has been done on FER, Identifying facial expressions in real life is difficult due to frequent movements of the head and subtle facial deformations, and other real-time variability that motivate researchers to provide their efforts to find a possible solution.

References

1. Abdullah, M., Wazzan, M., Bo-Saeed, S.: Optimizing face recognition using PCA. arXiv preprint arXiv:1206.1515 (2012)
2. Abdulrahman, M., Gwadabe, T.R., Abdu, F.J., Eleyan, A.: Gabor wavelet transform based facial expression recognition using PCA and LBP. In: 2014 22nd Signal Processing and Communications Applications Conference (SIU), pp. 2265–2268. IEEE (2014)

3. Aghamaleki, J.A., Chenarlogh, V.A.: Multi-stream CNN for facial expression recognition in limited training data. Multimedia Tools Appl. **78**(16), 22,861–22,882 (2019)
4. Alam, M., Vidyaratne, L.S., Iftekharuddin, K.M.: Sparse simultaneous recurrent deep learning for robust facial expression recognition. IEEE Trans. Neural Netw. Learn. Syst. **29**(10), 4905–4916 (2018)
5. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Trans. Pattern Anal. Mach. Intell. **19**(7), 711–720 (1997)
6. Butalia, M.A., Ingle, M., Kulkarni, P.: Facial expression recognition for security. Int. J. Mod. Eng. Res. **2**(4), 1449–1453 (2012)
7. Cai, J., Meng, Z., Khan, A.S., Li, Z., O'Reilly, J., Tong, Y.: Island loss for learning discriminative features in facial expression recognition. In: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), pp. 302–309. IEEE (2018)
8. Carrier, P.L., Courville, A., Goodfellow, I.J., Mirza, M., Bengio, Y.: FER-2013 face database. Universit de Montral (2013)
9. Cossetin, M.J., Nievola, J.C., Koerich, A.L.: Facial expression recognition using a pairwise feature selection and classification approach. In: 2016 International Joint Conference on Neural Networks (IJCNN), pp. 5149–5155. IEEE (2016)
10. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), vol. 1, pp. 886–893. IEEE (2005)
11. Dhall, A., Goecke, R., Lucey, S., Gedeon, T.: Static facial expressions in tough conditions: data, evaluation protocol and benchmark. In: 1st IEEE International Workshop on Benchmarking Facial Image Analysis Technologies BeFIT, ICCV2011 (2011)
12. Dureha, A.: An accurate algorithm for generating a music playlist based on facial expressions. Int. J. Comput. Appl. **100**(9), 33–39 (2014)
13. Ekman, P., Friesen, W.V.: Constants across cultures in the face and emotion. J. Pers. Soc. Psychol. **17**(2), 124 (1971)
14. Benitez-Quiroz, C.F., Srinivasan, R., Martinez, A.M.: EmotioNet: an accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5562–5570 (2016)
15. Fan, Y., Lu, X., Li, D., Liu, Y.: Video-based emotion recognition using CNN-RNN and C3D hybrid networks. In: Proceedings of the 18th ACM International Conference on Multimodal Interaction, pp. 445–450 (2016)
16. González-Lozoya, S.M., de la Calleja, J., Pellegrin, L., Escalante, H.J., Medina, M.A., Benitez-Ruiz, A.: Recognition of facial expressions based on CNN features. Multimedia Tools Appl. **79**(19), 13,987–14,007 (2020)
17. Gross, R., Matthews, I., Cohn, J., Kanade, T., Baker, S.: Multi-PIE. Image Vis. Comput. **28**(5), 807–813 (2010)
18. Gupta, S.K., Nain, N.: Gabor filter meanPCA feature extraction for gender recognition. In: Chaudhuri, B.B., Kankanhalli, M.S., Raman, B. (eds.) Proceedings of 2nd International Conference on Computer Vision & Image Processing. AISC, vol. 704, pp. 79–88. Springer, Singapore (2018). https://doi.org/10.1007/978-981-10-7898-9_7
19. Gupta, S.K., Agrwal, S., Meena, Y.K., Nain, N.: A hybrid method of feature extraction for facial expression recognition. In: 2011 Seventh International Conference on Signal Image Technology & Internet-Based Systems, pp. 422–425. IEEE (2011)

20. Hamester, D., Barros, P., Wermter, S.: Face expression recognition with a 2-channel convolutional neural network. In: 2015 International Joint Conference on Neural Networks (IJCNN), pp. 1–8. IEEE (2015)
21. Happy, S., Routray, A.: Automatic facial expression recognition using features of salient facial patches. IEEE Trans. Affect. Comput. **6**(1), 1–12 (2014)
22. Islam, B., Mahmud, F., Hossain, A., Goala, P.B., Mia, M.S.: A facial region segmentation based approach to recognize human emotion using fusion of hog & LBP features and artificial neural network. In: 2018 4th International Conference on Electrical Engineering and Information & Communication Technology (iCEEiCT), pp. 642–646. IEEE (2018)
23. Jabid, T., Kabir, M.H., Chae, O.: Facial expression recognition using local directional pattern (LDP). In: 2010 IEEE International Conference on Image Processing, pp. 1605–1608. IEEE (2010)
24. Jung, H., Lee, S., Yim, J., Park, S., Kim, J.: Joint fine-tuning in deep neural networks for facial expression recognition. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2983–2991 (2015)
25. Kamachi, M., Lyons, M., Gyoba, J.: The Japanese female facial expression (JAFFE) database (1998). http://www.kasrl.org/jaffe.html. 21:32
26. Kar, N.B., Babu, K.S., Jena, S.K.: Face expression recognition using histograms of oriented gradients with reduced features. In: Raman, B., Kumar, S., Roy, P.P., Sen, D. (eds.) Proceedings of International Conference on Computer Vision and Image Processing. AISC, vol. 460, pp. 209–219. Springer, Singapore (2017). https://doi.org/10.1007/978-981-10-2107-7_19
27. Kim, J.H., Kim, B.G., Roy, P.P., Jeong, D.M.: Efficient facial expression recognition algorithm based on hierarchical deep neural network structure. IEEE Access **7**, 41,273–41,285 (2019)
28. Kumar, S., Bhuyan, M.K., Chakraborty, B.K.: Extraction of informative regions of a face for facial expression recognition. IET Comput. Vision **10**(6), 567–576 (2016)
29. Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D.H., Hawk, S.T., Van Knippenberg, A.: Presentation and validation of the Radboud faces database. Cogn. Emot. **24**(8), 1377–1388 (2010)
30. Lee, J., Uddin, M.Z., Kim, T.S.: Spatiotemporal human facial expression recognition using fisher independent component analysis and hidden Markov model. In: 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 2546–2549. IEEE (2008)
31. Li, S., Deng, W.: Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition. IEEE Trans. Image Process. **28**(1), 356–370 (2018)
32. Li, Y., Zeng, J., Shan, S., Chen, X.: Occlusion aware facial expression recognition using CNN with attention mechanism. IEEE Trans. Image Process. **28**(5), 2439–2450 (2018)
33. Liu, M., Li, S., Shan, S., Wang, R., Chen, X.: Deeply learning deformable facial action parts model for dynamic expression analysis. In: Cremers, D., Reid, I., Saito, H., Yang, M.-H. (eds.) ACCV 2014. LNCS, vol. 9006, pp. 143–157. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-16817-3_10
34. Liu, M., Li, S., Shan, S., Chen, X.: Au-inspired deep networks for facial expression feature learning. Neurocomputing **159**, 126–136 (2015)
35. Liu, P., Han, S., Meng, Z., Tong, Y.: Facial expression recognition via a boosted deep belief network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1805–1812 (2014)

36. Liu, Y., Li, Y., Ma, X., Song, R.: Facial expression recognition with fusion features extracted from salient facial areas. Sensors **17**(4), 712 (2017)
37. Lucey, P., Cohn, J.F., 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 Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, pp. 94–101. IEEE (2010)
38. Mahmoudi, M.A., Chetouani, A., Boufera, F., Tabia, H.: Improved bilinear model for facial expression recognition. In: Djeddi, C., Kessentini, Y., Siddiqi, I., Jmaiel, M. (eds.) MedPRAI 2020. CCIS, vol. 1322, pp. 47–59. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-71804-6_4
39. Mandal, M.K., Pandey, R., Prasad, A.B.: Facial expressions of emotions and schizophrenia: a review. Schizophr. Bull. **24**(3), 399–412 (1998)
40. Mehrabian, A.: Communication without words. In: Communication Theory, pp. 193–200. Routledge (2017)
41. Mehta, N., Jadhav, S.: Facial emotion recognition using log Gabor filter and PCA. In: 2016 International Conference on Computing Communication Control and automation (ICCUBEA), pp. 1–5. IEEE (2016)
42. Mollahosseini, A., Chan, D., Mahoor, M.H.: Going deeper in facial expression recognition using deep neural networks. In: 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1–10. IEEE (2016)
43. Mollahosseini, A., Hasani, B., Mahoor, M.H.: AffectNet: a database for facial expression, valence, and arousal computing in the wild. IEEE Trans. Affect. Comput. **10**(1), 18–31 (2017)
44. Nazir, M., Jan, Z., Sajjad, M.: Facial expression recognition using histogram of oriented gradients based transformed features. Clust. Comput. **21**(1), 539–548 (2018)
45. Ng, H.W., Nguyen, V.D., Vonikakis, V., Winkler, S.: Deep learning for emotion recognition on small datasets using transfer learning. In: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, pp. 443–449 (2015)
46. Nigam, S., Singh, R., Misra, A.: Efficient facial expression recognition using histogram of oriented gradients in wavelet domain. Multimedia Tools Appl. **77**(21), 28,725–28,747 (2018)
47. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002)
48. Pantic, M., Valstar, M., Rademaker, R., Maat, L.: Web-based database for facial expression analysis. In: 2005 IEEE International Conference on Multimedia and Expo, p. 5. IEEE (2005)
49. Pramerdorfer, C., Kampel, M.: Facial expression recognition using convolutional neural networks: state of the art. arXiv preprint arXiv:1612.02903 (2016)
50. Ryu, B., Rivera, A.R., Kim, J., Chae, O.: Local directional ternary pattern for facial expression recognition. IEEE Trans. Image Process. **26**(12), 6006–6018 (2017)
51. Saurav, S., Gidde, P., Saini, R., Singh, S.: Dual integrated convolutional neural network for real-time facial expression recognition in the wild. Vis. Comput. **38**(3), 1083–1096 (2022)
52. Shan, C., Gong, S., McOwan, P.W.: Facial expression recognition based on local binary patterns: a comprehensive study. Image Vis. Comput. **27**(6), 803–816 (2009)
53. Shan, K., Guo, J., You, W., Lu, D., Bie, R.: Automatic facial expression recognition based on a deep convolutional-neural-network structure. In: 2017 IEEE 15th International Conference on Software Engineering Research, pp. 123–128. Management and Applications (SERA). IEEE (2017)

54. Soni, K., Gupta, S.K., Kumar, U., Agrwal, S.L.: A new Gabor wavelet transform feature extraction technique for ear biometric recognition. In: 2014 6th IEEE Power India International Conference (PIICON), pp. 1–3. IEEE (2014)
55. Susskind, J.M., Anderson, A.K., Hinton, G.E.: The Toronto face database. Technical Report, Department of Computer Science, University of Toronto, Toronto, Canada 3 (2010)
56. Ts, A., Guddeti, R.M.R.: Automatic detection of students' affective states in classroom environment using hybrid convolutional neural networks. Educ. Inf. Technol. **25**(2), 1387–1415 (2020)
57. Tsai, H.H., Chang, Y.C.: Facial expression recognition using a combination of multiple facial features and support vector machine. Soft. Comput. **22**(13), 4389–4405 (2018)
58. Turan, C., Lam, K.M., He, X.: Soft locality preserving map (SLPM) for facial expression recognition (2018). arXiv preprint arXiv:1801.03754
59. Uddin, M.Z., Lee, J., Kim, T.S.: An enhanced independent component-based human facial expression recognition from video. IEEE Trans. Consum. Electron. **55**(4), 2216–2224 (2009)
60. Uddin, M.Z., Hassan, M.M., Almogren, A., Zuair, M., Fortino, G., Torresen, J.: A facial expression recognition system using robust face features from depth videos and deep learning. Comput. Electric. Eng. **63**, 114–125 (2017)
61. Verma, K., Khunteta, A.: Facial expression recognition using Gabor filter and multi-layer artificial neural network. In: 2017 International Conference on Information, Communication, Instrumentation and Control (ICICIC), pp. 1–5. IEEE (2017)
62. Wang, K., Peng, X., Yang, J., Lu, S., Qiao, Y.: Suppressing uncertainties for large-scale facial expression recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6897–6906 (2020)
63. Wang, K., Peng, X., Yang, J., Meng, D., Qiao, Y.: Region attention networks for pose and occlusion robust facial expression recognition. IEEE Trans. Image Process. **29**, 4057–4069 (2020)
64. Wang, W., et al.: A fine-grained facial expression database for end-to-end multipose facial expression recognition (2019). arXiv preprint arXiv:1907.10838
65. Weldon, T.P., Higgins, W.E., Dunn, D.F.: Efficient Gabor filter design for texture segmentation. Pattern Recogn. **29**(12), 2005–2015 (1996)
66. Wen, G., Chang, T., Li, H., Jiang, L.: Dynamic objectives learning for facial expression recognition. IEEE Trans. Multimedia **22**(11), 2914–2925 (2020)
67. Wu, Y., Liu, W., Wang, J.: Application of emotional recognition in intelligent tutoring system. In: First International Workshop on Knowledge Discovery and Data Mining (WKDD 2008), pp. 449–452. IEEE (2008)
68. Xie, S., Hu, H.: Facial expression recognition using hierarchical features with deep comprehensive multipatches aggregation convolutional neural networks. IEEE Trans. Multimedia **21**(1), 211–220 (2018)
69. Xu, Q., Zhao, N.: A facial expression recognition algorithm based on CNN and LBP feature. In: 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), vol. 1, pp. 2304–2308. IEEE (2020)
70. Yadav, P., Poonia, A., Gupta, S.K., Agrwal, S.: Performance analysis of Gabor 2D PCA feature extraction for gender identification using face. In: 2017 2nd International Conference on Telecommunication and Networks (TEL-NET), pp. 1–5. IEEE (2017)
71. Yan, J., Zheng, W., Cui, Z., Tang, C., Zhang, T., Zong, Y.: Multi-cue fusion for emotion recognition in the wild. Neurocomputing **309**, 27–35 (2018)

72. Yang, H., Ciftci, U., Yin, L.: Facial expression recognition by de-expression residue learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2168–2177 (2018)
73. Yin, L., Wei, X., Sun, Y., Wang, J., Rosato, M.J.: A 3D facial expression database for facial behavior research. In: 7th International Conference on Automatic Face and Gesture Recognition (FGR06), pp. 211–216. IEEE (2006)
74. Yu, Z., Liu, Q., Liu, G.: Deeper cascaded peak-piloted network for weak expression recognition. Vis. Comput. **34**(12), 1691–1699 (2018)
75. Zeng, G., Zhou, J., Jia, X., Xie, W., Shen, L.: Hand-crafted feature guided deep learning for facial expression recognition. In: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), pp. 423–430. IEEE (2018)
76. Zhang, K., Huang, Y., Du, Y., Wang, L.: Facial expression recognition based on deep evolutional spatial-temporal networks. IEEE Trans. Image Process. **26**(9), 4193–4203 (2017)
77. Zhang, Z., Zhang, J.: A new real-time eye tracking for driver fatigue detection. In: 2006 6th International Conference on ITS Telecommunications, pp. 8–11. IEEE (2006)
78. Zhang, Z., Luo, P., Loy, C.C., Tang, X.: From facial expression recognition to interpersonal relation prediction. Int. J. Comput. Vision **126**(5), 550–569 (2018)
79. Zhao, Z., Liu, Q., Wang, S.: Learning deep global multi-scale and local attention features for facial expression recognition in the wild. IEEE Trans. Image Process. **30**, 6544–6556 (2021)