Pattern Analysis and Applications (2023) 26:381–402
https://doi.org/10.1007/s10044-022-01112-0

INDUSTRIAL AND COMMERCIAL APPLICATION

Fast facial expression recognition using Boosted Histogram of Oriented Gradient (BHOG) features

Sumeet Saurav[1,2] · **Ravi Saini**[1,2] · **Sanjay Singh**[1,2]

Received: 16 June 2021 / Accepted: 27 August 2022 / Published online: 20 September 2022

Abstract

Systems for automatic facial expression recognition (FER) have an enormous need in advanced human-computer interaction (HCI) and human-robot interaction (HRI) applications. Over the years, researchers developed many handcrafted feature descriptors for the FER task. These descriptors delivered good accuracy on publicly available FER benchmark datasets. However, these descriptors generate high dimensional features that increase the computational time of the classifiers. Also, a significant proportion of the features are irrelevant and do not provide additional information for facial expression analysis. Adversely, these redundant features degrade the classification accuracy of the FER algorithm. This study presents an alternate, simple, and efficient scheme for FER in static images using the Boosted Histogram of Oriented Gradient (BHOG) descriptor. The proposed BHOG descriptor employs the AdaBoost feature selection algorithm to select important facial features from the original high-dimensional Histogram of Oriented Gradient (HOG) features. The BHOG descriptor with a reduced feature dimension decreases the computational cost without diminishing the recognition accuracy. The proposed FER pipeline tuned on the optimal values of different hyperparameters achieves competitive recognition accuracy on five benchmark FER datasets, namely CK+, JAFFE, RaFD, TFE, and RAF-DB. Also, the cross-dataset experiments confirm the superior generalization performance of the proposed FER pipeline. Finally, the comparative analysis results with existing FER techniques revealed the effectiveness of the pipeline. The proposed FER scheme is computationally efficient and classifies facial expressions in real time.

Keywords Facial expression recognition · Feature selection · Histogram of oriented gradient (HOG) · AdaBoost feature selection · Kernel extreme learning machine (K-ELM) classifier

1 Introduction

Systems for automatic FER provide crucial cues that can reveal an individual's hidden intention and state of mind. Therefore, there has been a huge demand for robust and computationally efficient FER systems for numerous HCI-based advanced assistive technologies. Besides, these systems can play a vital role in security and surveillance applications. Therefore, over the years, researchers developed several methods for emotion analysis using different input modalities. These include the use of visual sensors (RGB [1], thermal [2], and depth [3]), audio sensors [4], and sensors that capture physiological signals such as electroencephalogram (EEG) [5], respiration (RSP), and heart rate variability (HRV). Techniques also exist for emotion recognition that uses combination of various physiological signals such as RSP and HRV [6], audio and visual [7], and EEG and visual signals [8, 9].

In the last few years, numerous works were proposed in the literature demonstrating the applicability of FER-related technology in all spheres of human life. Alhussein [10] proposed a FER technique for the initial assessment of patients in an e-Healthcare platform. The work presented by Jeong and Ko [11] has demonstrated the usefulness of FER technology in the Advanced Driver-Assistance

✉ Sumeet Saurav
sumeet@ceeri.res.in

Ravi Saini
ravi@ceeri.res.in

Sanjay Singh
sanjay@ceeri.res.in

[1] Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India

[2] CSIR-Central Electronics Engineering Research Institute (CSIR-CEERI), Pilani 333031, India

System (ADAS). Using EEG signals, Mehdizadehfar et al. [5] demonstrated the impact of facial emotion recognition in the fathers of children with autism. Recently, Sini et al. [12] confirmed the applicability of an automatic emotion recognition system in the calibration of autonomous driving functions. Li et al. [13] proposed a technique for FER and integrated it with a societal robot to enhance human-robot interaction (HRI). Finally, Yolcu et al. [14] proposed a FER-based system to detect facial expressions in people suffering from neurological disorders automatically.

Based on the learning scheme, the vision-based methods for FER can be divided broadly into two categories: the traditional machine learning-based approaches [15, 16], and deep learning-based approaches [17]. The machine learning-based approach for FER uses a combination of handcrafted feature extractors like the local binary pattern (LBP) and a machine learning classifier such as a support vector machine (SVM). Furthermore, since the dimensions of the handcrafted features are high, to overcome the curse of dimensionality and to reduce the overall computational time of the system, an optional dimensionality reduction or feature selection schemes have also been utilized in the FER task [15, 16, 18, 19]. Although the convolutional neural networks (CNNs), being data-driven, have attained state-of-the-art accuracy on several benchmark FER datasets, the FER methods based on traditional machine learning have also reported achieving competitive performance [20].

The current research in traditional machine learning-based FER domain has been towards designing discriminate and robust feature extractors [21], efficient feature selection algorithms [19], and powerful multi-class expression classifier [22]. This work investigates the effectiveness of the AdaBoost feature selection (FS) algorithm, Histogram of Oriented Gradient (HOG) feature extractor, and Kernel Extreme Learning Machine (K-ELM) classifier to implement a robust and computationally efficient system for FER. To this end, we proposed a FER pipeline that consists of four stages: input pre-processing, feature extraction, feature selection, and expression classification. For performance evaluation, the pipeline is validated on five FER benchmark datasets (CK+, JAFFE, RaFD, TFE, and RAF-DB) and compared with state-of-the-art FER methods. The main contributions of the proposed FER framework are as follows:

- Designed and implemented a computationally efficient and robust algorithmic pipeline for automatic FER using different HOG descriptors and AdaBoost feature selection (FS) algorithm.
- Deployment of the kernel extreme learning machine (K-ELM) classifier to classify several facial expressions. K-ELM has not been utilized much in FER tasks. However, it is computationally efficient compared to the popular classifiers such as the Support Vector Machine (SVM) and Naive Bayes (NB).
- Devised a set of procedures to choose the best values of various hyperparameters in the proposed FER pipeline.
- Performance analysis of the proposed FER pipeline using three testing procedures, namely the tenfold cross-validation, train-test evaluation, and cross-dataset testing, on five benchmark FER datasets: CK+, JAFFE, RaFD, TFE, and RAF-DB.

We organize the remaining contents of this paper into the following sections: Section 2 provides the details of the related FER works available in the literature. In Sect. 3, we provide the overview of the proposed FER pipeline, along with details of its constituent units. Section 4 provides details of the experimental setup and FER datasets. It also includes performance evaluation results on the datasets along with necessary discussions. Discussion on the computational performance analysis of the proposed and related FER methods makes the contents of Sect. 5. Finally, Sect. 6 concludes the study with conclusive comments and future research directions.

2 Related works

Based on the feature type, available techniques for static image-based FER are categorized into appearance feature-based methods, geometrical feature-based methods, and methods using the hybrid of the appearance and geometrical features [23]. Below, we briefly review the existing works on appearance feature-based methods for FER in static images.

2.1 Facial texture-based methods for FER

Over the years, researchers developed several texture descriptors for image classification tasks. Among these descriptors, the Local Binary Pattern (LBP) and its variants have been utilized widely in the FER-related work [15, 24–27]. Other advanced texture descriptors proposed for the FER task include the Gradient Local Ternary Pattern (GLTP) and its improved variant [16, 28], Improved Completed Local Ternary Patterns (ICLTP) [29], Neighborhood-aware Edge Directional Pattern (NEDP) [30], Dynamic Local Ternary Pattern (DLTP) [31], Gradient Local Phase Quantization (GLPQ) [32], Local Directional Ternary Pattern (LDTP) [33], and so on. Alhussein [10] introduced the multi-scale variant of the Weber Local Descriptor (MS-WLD) for FER. Working on a similar line, other researchers developed improved variants of WLD, such as the Weber Local Binary Image Cosine Transform (WLBI-CT) [34] and DCT transformed WLD descriptor [35]. The FER method proposed by Siddiqi et al. [36] has used curvelet transform to extract features from the facial images. The Elongated Quinary Pattern (EQP) with five-level encoding utilized by Al-Sumaidaee et al. [37] extracts highly discriminate facial features from Sobel convolved gradient magnitude and angular facial images.

Recently, Alphonse and Starvin [38] introduced two new directional patterns named the Maximum Response-based Directional Texture Pattern (MRDTP) and the Maximum Response-based Directional Number Pattern (MRDNP) for FER in constrained and unconstrained scenarios. Facial features extracted using MRDTP and MRDNP were first reduced using the Generalized Supervised Dimension reduction system (GSDRS) and eventually classified using the Extreme Learning Machine with Radial Basis Function (ELM-RBF) classifier. Gogić et al. [39] proposed a fast and efficient pipeline for FER that uses Local Binary Features (LBF) descriptor to extract features from the facial images and a shallow neural network (NN) to classify the features into different expressions. Revina and Emmanuel [22] utilized the combination of the Scale-Invariant Feature Transform (SIFT) and a new texture descriptor called Scatter Local Directional Pattern (SLDP) for the FER task. The features derived from the facial images were classified using the Multi-Support Vector Neural Network classifier optimized using the Whale-Grasshopper Optimization algorithm. In their other work [40], the authors employed the Support Vector Neural Network (SVNN) classifier to classify facial image features extracted using the Multi-Directional Triangles Pattern (MDTP) descriptor.

2.2 Facial shape-based methods for FER

Facial expression changes the shape of facial muscles, and thus, descriptors that encode these changes in facial muscles may be useful in identifying the expressions. Subsequently, researchers developed several descriptors to extract shape-based facial information corresponding to different facial expressions. Carcagnì et al. [1] conducted a detailed study to analyze the impact of the HOG hyperparameters (cell size, number of bins, and type of orientations) on the recognition accuracy of the FER system. In other work [41], instead of directly using the HOG extracted facial features, the authors suggested using the difference of the features derived from the neutral and peak expression images. Additionally, the authors utilized the genetic algorithm to determine the optimal values of the HOG parameters. The FER scheme introduced by Nazir et al. [42] has also utilized the HOG descriptor for FER in facial images of different resolutions (128×128, 64×64, 32×32). Nigam et al. [43], on the other hand, proposed an advanced variant of the HOG descriptor named W_HOG for the FER task. The W_HOG descriptor, as the name suggests, applies HOG on the discrete wavelet transformed (DWT) facial images.

2.3 Hybrid methods for FER

Several FER techniques in the literature have also applied the fusion of facial texture and shape information extracted using appearance and shape descriptors, respectively. For instance, the multiple kernel learning (MKL) based FER scheme proposed by Zhang et al. [44] has utilized a fusion of HOG and LBP extracted facial features. Working on a similar line, Liu et al. [20] suggested fusing LBP and HOG features extracted from silent facial regions. The work presented in [45] has analyzed the effectiveness of facial texture features, facial shape features, and a hybrid of facial texture and shape features for the FER task. Yang et al. [46] suggested fusing the facial appearance features extracted using LBP and deep geometric features extracted using facial landmarks. Their proposed FER pipeline utilized the Random Forest (RF) classifier to classify the fused features into six basic facial expressions. The FER scheme introduced by Ghimire et al. [47] has employed the hybrid of appearance and geometrical features. Once extracted, the appearance and geometric features extracted from domain-specific facial regions were concatenated and classified using the SVM classifier. In their recent work, Shanthi and Nickolas [48] proposed a fusion of features derived from the facial images using LBP and the newly introduced local neighborhood encoded pattern (LNEP).

2.4 Feature selection techniques for FER

Besides feature descriptors, researchers employed/developed several feature selection (FS) algorithms for the FER task in the last decade. These algorithms aim to improve the computational efficiency of the FER system without much degradation in its classification accuracy. The FS algorithms, by reducing the dimension of the facial features, not only enhance the classification time but, in most cases, also leads to improvement in the classification accuracy. Lajevardi and Hussain [49] investigated the role of the minimum redundancy–maximum relevance (mRMR) feature selection (FS) algorithm for the FER task. In their other work [50], the authors analyzed the effectiveness of the mutual information feature selection (MIFS) algorithm, the mutual information quotient (MIQ) algorithm, and the Genetic algorithm (GA) for facial feature selection. Siddiqi et al. [36] introduced a normalized mutual information-based FS technique for FER that normalizes the mutual information and reduces the dominance of relevance or redundancy. For information regarding recent FS techniques for the FER task, we refer the readers to some of the recent works [18, 19].

Based on the above review, one can find that despite tremendous advancement, the existing algorithms for static image-based FER using traditional machine learning are not suitable for deployment in real-world conditions. There is a huge requirement for robust and compute-efficient FER algorithms for real-world applications. This work is an attempt to improve the robustness and computational efficiency of the FER algorithm.

3 Proposed FER pipeline

The proposed FER pipeline shown in Fig. 1 involves five units executed in the following sequence: (1) face detection & landmark localization, (2) facial alignment & registration, (3) feature extraction, (4) feature selection, and (5) feature classification. The face detection & landmark localization unit determines the locations of faces and facial landmarks in the input frame. Based on the locations of the faces and landmarks in the input image, the face alignment & registration unit provides aligned facial images of a standard size. The feature extraction unit using the HOG descriptor extracts features from the registered facial images. The dimensions of the HOG extracted features are high, and they include many redundant features. Therefore, the HOG features are passed to the AdaBoost feature selection (FS) algorithm to reduce their dimension and select only the relevant facial features. Intuitively, out of many, the feature selector selects the expression-specific active facial patches (see Fig. 1). Finally, the feature classification unit using the Kernel Extreme Learning Machine (K-ELM) classifier classifies the selected features into facial expressions. Below, we provide further details of these constituent units.

3.1 Face detection and landmark localization

The face detection & landmark localization unit utilizes the Viola & Jones face detector [51] and Intraface facial landmark detector [52] to obtain the *x* and *y* coordinates of the face and the 68-facial landmarks, respectively. The face detector uses the cascade classifier trained on Multi-block local binary pattern (MB-LBP) [53]. Once detected, the unit passes face coordinates to the facial landmark localizer. The localizer uses the Supervised Descent Method (SDM) to mark the location of 68-facial landmarks on the detected faces.

3.2 Face alignment and registration

The face alignment & registration unit utilize the face and facial landmarks coordinates for facial image registration. Essentially, the unit uses the landmark coordinates of the eyes to compute the inter-ocular distance (D) and inter-angle between the eyes' center. In the subsequent step, the unit affine transforms the facial images for rotation rectification and crops the face region using the predefined value of *D*, as shown in Fig. 2. The cropping scheme discards redundant face regions and ensures spatial symmetry of facial components [1, 16]. Finally, the face alignment & registration unit scale the cropped image to a standard resolution.

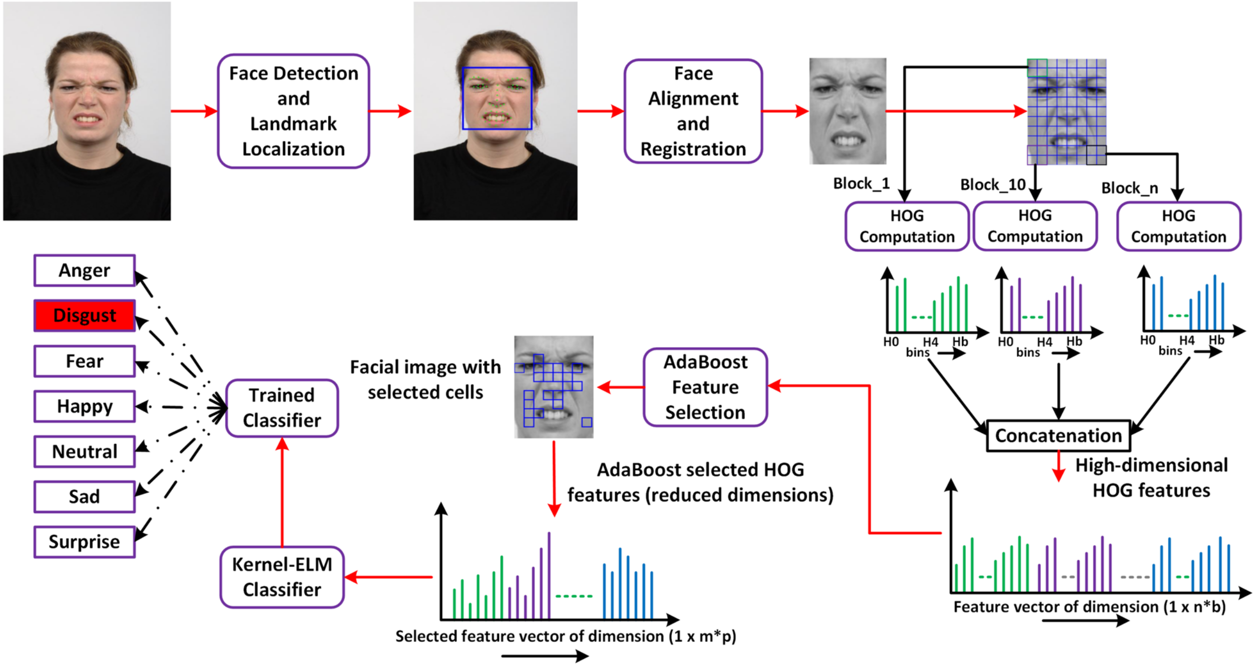
Face Detection and Landmark Localization
Face Alignment and Registration
Block_1
Block_10
Block_n
HOG Computation
HOG Computation
HOG Computation
H0 H4 Hb
bins
H0 H4 Hb
bins
H0 H4 Hb
bins
Concatenation
High-dimensional HOG features
Feature vector of dimension (1 x n*b)
AdaBoost Feature Selection
Facial image with selected cells
AdaBoost selected HOG features (reduced dimensions)
Selected feature vector of dimension (1 x m*p)
Kernel-ELM Classifier
Trained Classifier
Anger
Disgust
Fear
Happy
Neutral
Sad
Surprise

Fig. 1 Algorithmic pipeline of the proposed facial expression recognition system

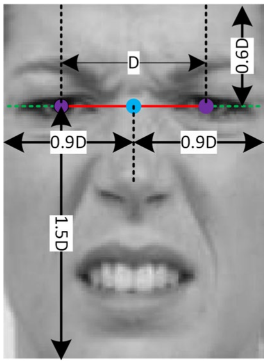

Fig. 2 Scheme used to crop facial image

3.3 Feature extraction

This study has employed three variants of the popular Histogram of Oriented Gradient (HOG) descriptor for feature extraction from the facial images. The first variant, named HOGu, uses unsigned orientations, while the second variant, referred to as HOGs, using signed orientations, helps differentiate light-to-dark versus dark-to-light transitions in the facial images. The final variant termed HOGv is more compact and efficient. It uses both signed and unsigned orientations to extract enhanced expression details from the facial images [54]. The gradient orientation bins in HOGu are evenly spaced between 0° to +180°, whereas in the HOGs variant, the bins are evenly divided between −180° to +180°. The descriptor places theta values less than 0° into the theta + 180° value bin.

The HOG descriptor has several hyperparameters (cell size, block size, number of orientation bins, and block overlapping). The optimal values of these hyperparameters make HOG one of the most efficient descriptors that can extract discriminative features from facial images. Figure 3 shows the systematic representation of the steps involved in the extraction of facial features by the HOG descriptor. These steps are divided into: (1) Gradient magnitude and angle computation, (2) Gradient voting, and (3) Histogram normalization.

3.3.1 Gradient magnitude and angle computation

As shown in Fig. 3, each pixel $I(x, y)$ of a cell is convolved with predefined horizontal and vertical filters to generate horizontal gradient $G_x(x, y)$ and vertical gradient $G_y(x, y)$, respectively. The gradient calculation can be expressed mathematically as in Eqs. (1) and (2).

$$G_x(x, y) = I(x + 1, y) - I(x - 1, y) \tag{1}$$

$$G_y(x, y) = I(x, y + 1) - I(x, y - 1) \tag{2}$$

In the subsequent step, using $G_x(x, y)$ and $G_y(x, y)$, the gradient magnitude $G(x, y)$ and orientation $\theta(x, y)$ is computed using Eqs. (3) and (4), respectively.

$$G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2} \tag{3}$$

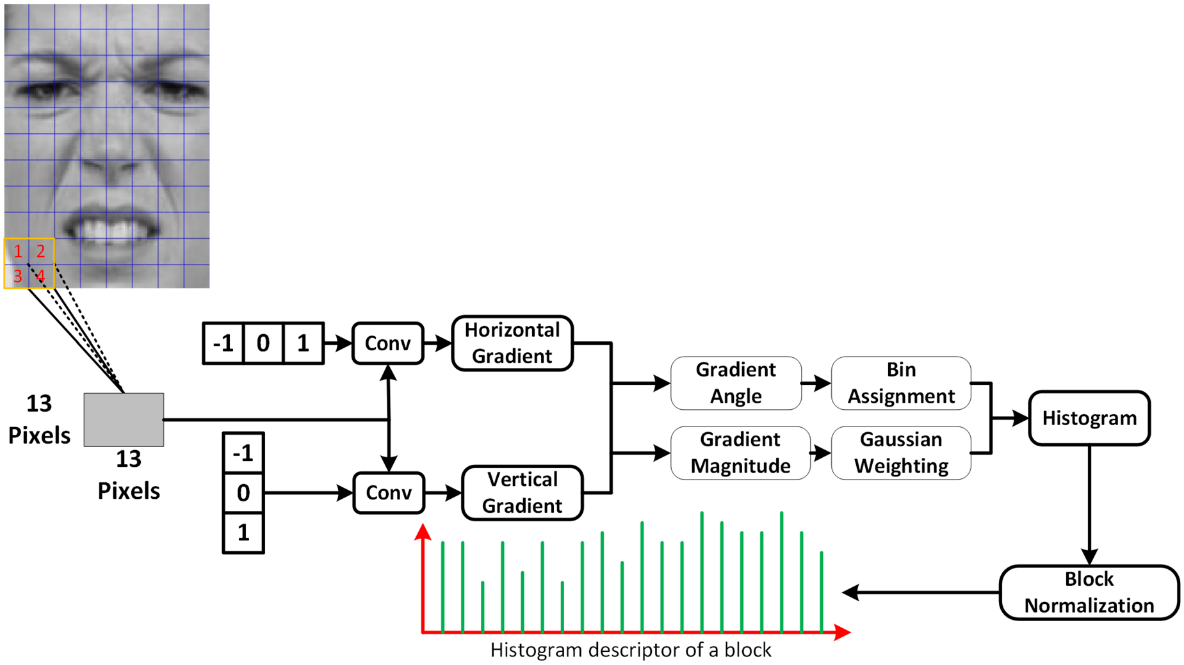

Fig. 3 Sequence of steps used for feature extraction by the HOG descriptor

$$\theta(x, y) = \arctan \frac{G_y(x, y)}{G_x(x, y)} \tag{4}$$

Once the gradient magnitude of all the cells in a block has been calculated, the combined block magnitude is then multiplied with a Gaussian matrix f_g of kernel size equal to the size of the block to get Gaussian weighted magnitude G_g [55], as illustrated in Eq. (5). Finally, the weighted gradient magnitudes are used for gradient voting, as discussed below.

$$G_g(x, y) = f_g * G(x, y) \tag{5}$$

3.3.2 Gradient voting

Each pixel within a cell donates a weighted vote in favor of an orientation histogram based on its gradient magnitude. The gradient magnitude corresponding to each pixel in the cell, in turn, is multiplied by a weight factor denoted by α, as discussed in [56] and computed using Eq. (6).

$$\alpha = (n + 0.5) - \frac{b * \theta(x, y)}{\pi} \tag{6}$$

In Eq. (6), n is the bin to which the gradient orientation $\theta(x, y)$ belongs, and b is the value of the orientation bins. To overcome the aliasing artifacts, the histogram values of neighborhood bins, G_n and $G_{nearest}$ are multiplied by weight factors (1 - α) and α, as illustrated in Eqs. (7) and (8), respectively. It is worth mentioning that in Eqs. (7) and (8), the Gaussian weighted magnitude G_g is used only in the case of HOGu and HOGs variants. For the HOGv variant, the operations are performed on the gradient magnitude G.

$$G_n = (1 - \alpha) * G_g(x, y) \tag{7}$$

$$G_{nearest} = \alpha * G_g(x, y) \tag{8}$$

3.3.3 Histogram normalization

As discussed in the previous step, in the gradient voting step, each pixel in the cell voted in favor of an orientation histogram bin based on its gradient magnitude. Let **S**(s, t) denote the histogram having semicircle orientation with b bins in the range from 0° to +180°. All three variants of the HOG descriptor have eventually utilized such a histogram. Besides, the HOGv variant uses an additional histogram denoted as **C**(s, t). It corresponds to circular gradient orientations with b equally divided orientation bins ranging from 0° to 360°. Thus, the HOGu and HOGs variants only use contrast-sensitive orientations, whereas the HOGv variant includes both contrast-sensitive and contrast-insensitive gradient orientations. For each cell histogram, the normalization operation proceeds in two steps [57]. In the first step, gradient energy is computed for the block containing the cells. For the histograms **S**(s, t) and **C**(s, t) of the cell indexed by (s, t), the gradient energy $HS_{\delta,\gamma}(s, t)$ and $HC_{\delta,\gamma}(s, t)$ are estimated using Eqs. (9) and (10), respectively, where, the variables $\delta, \gamma \in \{$-1,1$\}$.

$$\begin{aligned} HS_{\delta,\gamma}(s, t) = \big[&\|\mathbf{S}(s, t)\|^2 + \|\mathbf{S}(s + \delta, t)\|^2 \\ &+ \|\mathbf{S}(s, t + \gamma)\|^2 + \|\mathbf{S}(s + \delta, t + \gamma)\|^2 \big]^{\frac{1}{2}} \end{aligned} \tag{9}$$

$$\begin{aligned} HC_{\delta,\gamma}(s, t) = \big[&\|\mathbf{C}(s, t)\|^2 + \|\mathbf{C}(s + \delta, t)\|^2 \\ &+ \|\mathbf{C}(s, t + \gamma)\|^2 + \|\mathbf{C}(s + \delta, t + \gamma)\|^2 \big]^{\frac{1}{2}} \end{aligned} \tag{10}$$

The gradient energies are subsequently utilized in the second step to normalize the histogram of the cell, as demonstrates in Eq. (11).

$$\mathbf{F}(s,t) = \begin{pmatrix} \mathbf{S}(s, t)/HS_{-1,-1}(s, t), \mathbf{C}(s, t)/HC_{-1,-1}(s, t) \\ \mathbf{S}(s, t)/HS_{+1,-1}(s, t), \mathbf{C}(s, t)/HC_{+1,-1}(s, t) \\ \mathbf{S}(s, t)/HS_{+1,+1}(s, t), \mathbf{C}(s, t)/HC_{+1,+1}(s, t) \\ \mathbf{S}(s, t)/HS_{-1,+1}(s, t), \mathbf{C}(s, t)/HC_{-1,+1}(s, t) \end{pmatrix} \tag{11}$$

3.3.4 Feature extraction using HOGu and HOGs

Figure 4 shows feature extraction scheme using the HOGu and HOGs descriptors. The input facial image of dimension 147 × 108 is divided into 70 cells of 13 × 13 pixels. The cells are grouped into several blocks, wherein each block consists of 2 × 2 cells, each from the horizontal and vertical directions, with an overlap of one in both directions. In the case of HOGu, the gradient orientations in the range from 0° to +180° are divided evenly into 21 histogram bins. While in the case of HOGs, the gradient orientations in the range from −180° to +180° are first converted in the range from 0° to +180° and then divided into 21 equally spaced histogram bins. Thus, from each of the four cells of a block, the descriptors extract 1 × 21-dimensional cell histogram. Histograms from the four cells of a block are concatenated and normalized to obtain 1 × 84-dimensional block histogram. Finally, all block histograms are concatenated to obtain the final 1 × 5880(=70 × 84) dimensional facial feature **x**.

3.3.5 Feature extraction using HOGv

Figure 5 shows details of the feature extraction framework using the HOGv descriptor. The descriptor divides the input facial image into 88 equal-sized cells of 13 × 13 pixels. Four cells, two from horizontal and vertical directions, are grouped into blocks with an overlap of one in these directions. Also, except for the boundary cells, other cells of the

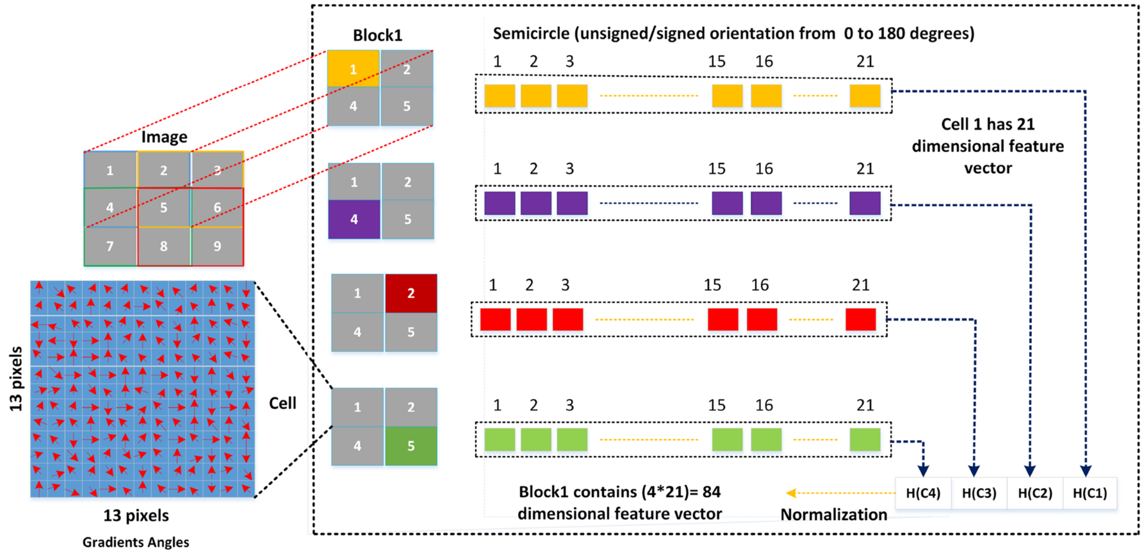
Block1
Semicircle (unsigned/signed orientation from 0 to 180 degrees)
1 2 3 15 16 21
Cell 1 has 21 dimensional feature vector
Image
1 2 3 4 5 6 7 8 9
13 pixels
Cell
13 pixels
Gradients Angles
Block1 contains (4*21)= 84 dimensional feature vector
Normalization
H(C4) H(C3) H(C2) H(C1)

Fig. 4 Feature extraction using HOGu and HOGs descriptors

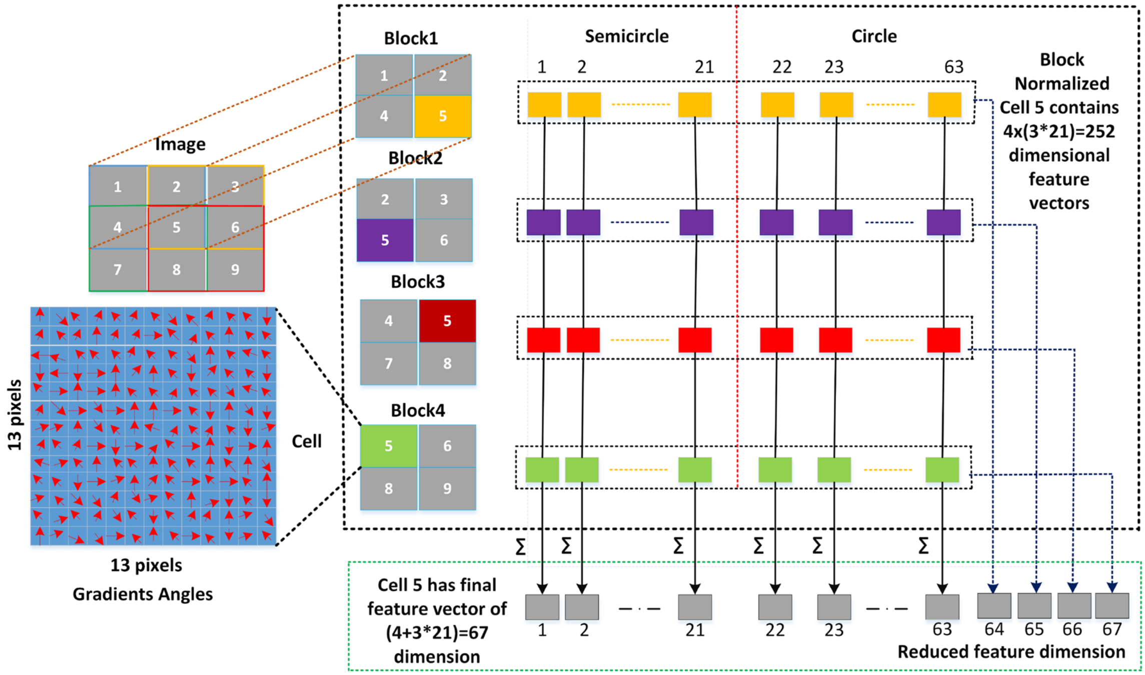
Block1
Block2
Block3
Block4
Semicircle
Circle
1 2 21 22 23 63
Block Normalized Cell 5 contains 4x(3*21)=252 dimensional feature vectors
Image
1 2 3 4 5 6 7 8 9
13 pixels
Cell
13 pixels
Gradients Angles
Σ Σ Σ Σ Σ Σ
Cell 5 has final feature vector of (4+3*21)=67 dimension
1 2 21 22 23 63 64 65 66 67
Reduced feature dimension

Fig. 5 Feature extraction using HOGv descriptor

image are shared by four neighborhood blocks. Therefore, for each cell shared by four blocks, a 1 × 252(=4 × 1 × 63)) dimensional features are extracted in the form of histogram bins corresponding to both semicircular and circular orientations.

In contrast to the HOGs and HOGu variant, the HOGv variant uses a principal component analysis-like scheme to reduce the dimensions of the features. As shown in Fig. 5, the dimensionality reduction scheme uses column-wise and row-wise summations of bins of the cell to help the descriptor capture the overall gradient energy from all the four neighborhood blocks. Since each cell has 21 bins for the semicircular orientations and 42 bins for the circular orientations, 63 column summations are performed, contributing 1 × 63 features to the final cell features. Meanwhile, the row summation captures the gradient energy of the cell over 63 orientations for each of the four neighborhood blocks, contributing an additional 1 × 4 features to the overall cell features. Therefore, the final cell-based feature vector has a dimension of 1 × 67 compared to the original 1 × 252-dimensional feature vector. In the final step of computation, the descriptor concatenates the histogram features extracted from all the cells in the image to obtain the final feature vector **x** of dimension 1 × 5896(=88 × 67).

3.4 Feature selection

In image classification tasks, features play a fundamental role in successful learning. Usually, the dimension of the features obtained after the feature extraction is high. Therefore, researchers developed several feature selection (FS) techniques to reduce the dimensions of the features. The FS algorithms select a subset of features from the original feature set such that the performance of the classifier, when trained on the selected subset of features, is at least equal or better than the performance obtained with the classifier trained on the original feature set. Thus, the FS methods provide a way of reducing the computational complexity of the classifier besides improving the classification performance. Besides, these techniques provide a better understanding of the data in machine learning or pattern recognition applications [58].

Based on their operations, the available techniques for FS are classified broadly into three categories: filter methods, wrapper methods, and embedded methods [59]. This study has used the boosting-based wrapper method called AdaBoost to select vital features from the original high-dimensional HOG features [60]. AdaBoost, an important meta-algorithm originally proposed for the pattern classification task, has also been utilized for FS in the FER tasks [15]. The AdaBoost algorithm utilizes a simple decision stump, a kind of decision tree with only one node, as the base learner. During training, the algorithm learns to find a weighted combination of decision stumps and use the combination as a strong and efficient classifier. The simple decision stump classifier used by AdaBoost for feature classification might not be very accurate. However, from the feature selection perspective, its performance is sufficient, as a trained decision stump corresponds to a selected feature.

Algorithm 1 illustrates computation steps employed by the AdaBoost FS algorithm. Input to the algorithm is facial features with corresponding expression labels, the number of iterations T, and the minimum accepted error Δ_{min}. The algorithm terminates once it reaches the specified Δ_{min} or the iterations T. Since the naive implementation of the AdaBoost algorithm is primarily binary, we used the one-versus-rest (OVR) multi-class scheme to select features from seven basic facial expressions. Feature selection using the OVR runs the binary AdaBoost FS algorithm seven times (equal to the number of facial expressions), and each run uses features from one of the facial expressions as positive class and the remaining expression as negative class. Finally, the FS scheme concatenates the features selected in each run to obtain the final subset of AdaBoost selected HOG features.

Algorithm 1: AdaBoost FS algorithm

```
Input: $(x_1, y_1)$,...., $(x_m, y_m)$, T, $\Delta_{min}$
initialize: $D_1(i) = \frac{1}{m}$, $SF \leftarrow \emptyset$
for $t$=$1$ to $T$ do
    Train stump $h_t(\mathbf{x})$ over $D_t(i)$; $h_t(\mathbf{x}) = \arg\min$
     $\epsilon_t$;
    $\epsilon_t = \frac{1}{2}\,[1 - \sum_i D_t(i) y_i h_t(x_i)]$
    if $\epsilon_t \geq \frac{1}{2}$ or $\epsilon_t \leq \Delta_{min}$ then
        stop
    end
    if Node of $h_t(\boldsymbol{x}) \notin SF$ then
        SF $\leftarrow h_t(\mathbf{x})$
        Calculate $\alpha_t = \frac{1}{2} \log \frac{1-\epsilon_t}{\epsilon_t}$
        $D_{t+1}(i) = \frac{D_t(i) \exp(-\alpha_t y_i h_t(\mathbf{x}))}{Z_t}$
    end
end
Output: SF;
```

3.5 Kernel extreme learning machine (K-ELM) classifier

The proposed FER pipeline has utilized the kernelized version of the Extreme Learning Machine (ELM) [61] classifier named the Kernel ELM (K-ELM) for the classification of facial expressions. Although the ELM classifier works very well in most pattern classification applications, the K-ELM classifier is optimal in situations where the feature transformation function is unknown [57]. Also, the classical ELM classifier requires many hidden nodes to achieve better performance. A large number of hidden nodes increases the computational cost of the classifier and its training time. It also increases the sensitiveness in the classifiers' performance due to the randomness of the parameters [62]. While the kernels in the K-ELM classifier directly map the features into higher dimensional space. Thus, apart from enhancing intra-class separability among the facial expressions, it also attains stable performance. For a detailed discussion on the ELM and the K-ELM classifier, we refer the readers to recent work on FER [31].

4 Experimental results and discussions

This section discusses the experimental details performed to obtain the best values of the facial image & the cell size, the number of the AdaBoost selected features, and the values of the parameters of the K-ELM classifier. We conducted these experiments in the MATLAB 2015a environment running on a Windows 10 machine with 16 GB RAM. For feature extraction using the HOGs and HOGu descriptor, we used the HOG implementation that comes with MATLAB. For feature extraction using the HOGv descriptor, on the other hand, the experiments used the MATLAB executable code of the descriptor that comes with the open-source VLFeat computer vision library [63].

4.1 Database details

In this section, we provide details of the CK+, JAFFE, RaFD, TFE, and RAF-DB datasets used in this study to investigate the effectiveness of the proposed FER pipeline.

4.1.1 CK+

The CK+ dataset is an extended variant of the Cohn-Kanade dataset and has expression sequences of both male and female participants [64]. The expression sequences contain images that start with the neutral expression and end at the peak expression. For a fair comparison, we prepared the dataset following the standard protocol used in the static image-based FER [39]. The final dataset contains 1236 facial images belonging to seven expressions having distribution as adopted by Saurav et al. [31] in their work on FER.

4.1.2 JAFFE

The Japanese female facial Expression (JAFFE) dataset is another in-the-lab FER dataset collected with the participation of ten Japanese female actresses [65]. The dataset contains 213 expressive facial images belonging to seven facial expressions: anger, disgust, fear, happiness, neutrality, sadness, and surprise.

4.1.3 RaFD

The Radboud Faces Database (RaFD) is a new FER dataset introduced to validate the performance of FER algorithms in static images [66]. Sixty-seven participants posed for the expressions with three gaze directions and five facial orientations during the dataset preparation. In this study, for a fair comparison with the existing works, we created five sub-categories (RaFD Category-1, RaFD Category-2, RaFD Category-3, RaFD Category-4, and RaFD Category-5) from the original RaFD dataset. All the five sub-categories of the dataset contain only the frontal facial images with different gaze directions. The RaFD Category-1 dataset consists of 469 (=67×7) frontal gaze facial images belonging to anger, contempt, disgust, fear, happiness, sadness, and surprise expressions. The second category of the dataset, named RaFD Category-2, consists of seven prototypical facial expressions (anger, disgust, fear, happiness, neutral, sadness, and surprise) having a distribution similar to the RaFD Category-1. The RaFD Category-3 dataset has 536 (=67×8) frontal gaze facial images from anger, contempt, disgust, fear, happiness, neutral, sad, and surprise expressions. While the fourth category of the dataset, named the RaFD Category-4, is the extended version of the Category-2 dataset and has facial images from all three gaze directions (left looking, right looking, and frontal), making a total of 1407 (=201×7) images in the dataset. Similarly, the RaFD Category-5 dataset is the extended variant of the Category-3 dataset, and it contains 1608 (=201×8) facial images from all three gaze directions (left looking, right looking, and frontal).

4.1.4 TFE

The Tsinghua facial expression (TFE) dataset is a recently introduced FER dataset that consists of facial images belonging to eight facial expressions: anger, contemptuous, disgust, fear, happiness, neutral, sadness, and surprise [67]. The dataset is the first of its kind introduced to study age-associated changes in facial expressions in the lab conditions and has images captured by 110 (63 young and 47 old) Chinese male and female adults. In this study, on the TFE dataset, we conducted three experiments to analyze the efficiency of the proposed FER pipeline. The age-independent FER experiment splits the complete dataset into train and test set in the 2:1 ratio. Out of the 63 young and 47 old subjects, the train set contains 583 facial images belonging to 42 young and 31 old subjects, while the test set contains 295 facial images belonging to the rest of the subjects. In the first age-dependent experiment, the train set consists of 376 facial images of the old subjects, and the test set contains 502 facial images of young subjects. In contrast, the age-dependent second experiment uses 502 facial images of the young and 376 facial images of the old subjects as the train and test set, respectively.

4.1.5 RAF-DB

RAF-DB is a real-world FER dataset that contains 30,000 facial images labeled with six basic expressions (anger, disgust, fear, happy, sad, and surprise) plus neutral and twelve compound expressions [68]. Forty trained independent labelers annotated each of the facial images in the dataset. Our experiment used only facial images with basic expressions, including 12,271 facial images as the training set and 3,068 facial images as the test set.

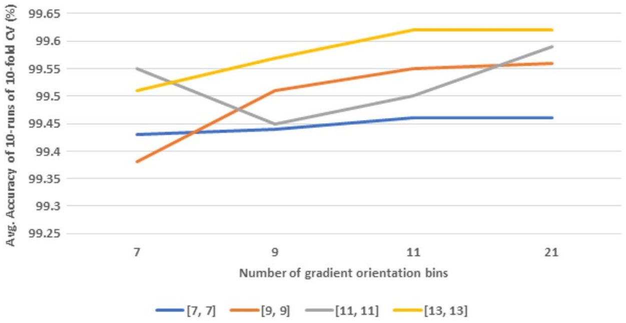

Fig. 6 Accuracy curves using different combinations of cell size and the number of histogram bins using the HOGu features extracted from the CK+ dataset

4.2 Determination of optimal values of hyperparameters

As discussed in Sections 3.3, 3.4, and 3.5 , the proposed FER pipeline has several hyperparameters. Therefore, the initial experiments were conducted on the CK+ dataset to determine the cell size, the number of histogram bins, and AdaBoost weak learners (number of AdaBoost selected features). The rest of the four FER datasets use the same optimal values of these hyperparameters. However, to determine the regularization coefficient (C) and kernel parameter (σ) of the K-ELM classifier, experiments were performed separately for all the five FER datasets.

4.2.1 Cell size and the number of histogram bins

The initial set of experiments conducted to determine the optimal value of the cell size and histogram orientation bins divides the facial images into multiple fixed-size cells (7×7, 9×9, 11×11, and 13×13) and use different orientation bins (7, 9, 11, and 21). From the facial images of the CK+ dataset, the proposed FER pipeline extracts block-wise HOGs features using all possible combinations of the cell sizes and orientation bins. The extracted features are classified by the K-ELM classifier using the fixed value of regularization parameter C=100 and kernel parameter σ=200.

Examining the recognition accuracy curves of Fig. 6, one can find that using the cell size of 13×13-pixels with the number of orientation bins set equal to 21, the HOGs extracted facial features achieved the best performance. The remaining experiments using the other two HOG variants on the rest of the FER datasets employed the same cell size and orientation bin values. Figure 7 shows HOG processed facial images obtained using the optimal value of the two hyperparameters on sample facial images from the CK+ dataset. Different variants of the HOG descriptor, using the values of cell size set to 13×13-pixels and orientation bins equal to 21, can efficiently extract the expression-related shape information from the facial images.

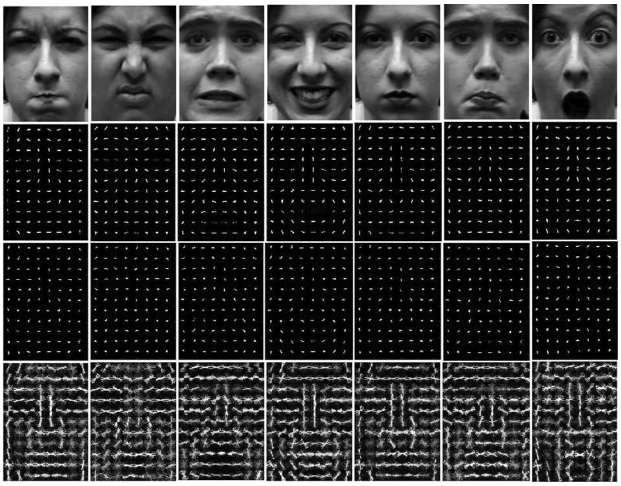

Fig. 7 HOG processed images obtained using optimal values of cell size and orientation bins on the CK+ dataset (top to bottom): Original facial images, HOGs processed facial images, HOGu processed facial images, and HOGv processed facial images (left to right): Anger, disgust, fear, happy, neutral, sad, and surprise

4.2.2 Optimal number of selected features

Figures 8, 9, and 10 show the accuracy curves of experiments conducted to determine the optimal number of AdaBoost weak learners or features per expressions. These experiments have used the original high-dimensional HOGs, HOGu, and HOGv features and the multi-class variant of the AdaBoost FS algorithm to select features from the seven facial expressions.

These experiments were performed on the CK+ dataset, varying the number of weak learners or features per expression from 25 to 150 at a regular interval of 25. We fixed the values of the rest of the hyperparameters ($\Delta_{min} = 0.01$, C = 100, and σ= 200) during the experiments. Analyzing the results of the experiments (see Figs. 8, 9, and 10), one can find that the number of weak learners or features per expression that resulted in the optimal performance was 125 for all three variants of the HOG descriptor. Also, the cumulative sum of the features selected from all the seven expressions were 276, 303, and 294 for the HOGs, HOGu, and HOGv descriptor, respectively. As expected, there were many features common for all the expressions. Also, on the CK+ dataset, the ten runs of 10-fold CV using the AdaBoost selected HOGs, HOGu, and HOGv feature achieved competitive mean recognition accuracy of 98.49%, 98.52%, and 98.49%, respectively. Figure 11 shows that the AdaBoost FS algorithm successfully selected the relevant action units (AUs) for each of the seven facial expressions: anger,

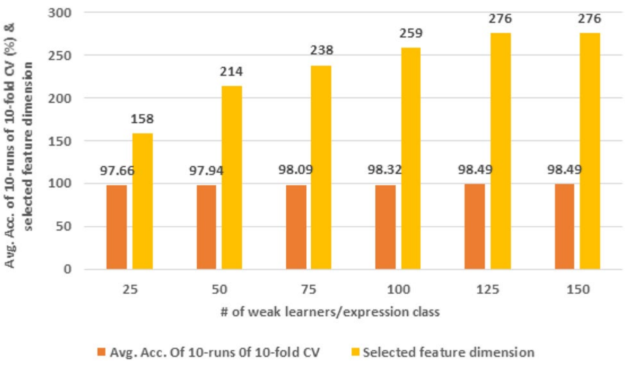

Fig. 8 Accuracy curves obtained by varying the number of weak learners or features per expression using AdaBoost FS algorithm and HOGs features extracted from the CK+ dataset

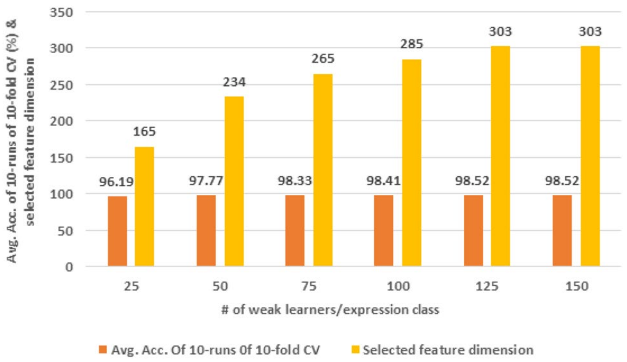

Fig. 9 Accuracy curves obtained by varying the number of weak learners or features per expression using AdaBoost FS algorithm and HOGu features extracted from the CK+ dataset

disgust, fear, happy, neutral, sad, and surprise, using all the three variants of the HOG descriptor. Also, as anticipated, the algorithm selected a comparatively larger number of cells for the neutral expression than other facial expressions. Thus, in the BHOG features, a large fraction of features are from the neutral class compared to other six basic expressions.

4.2.3 K-ELM classifier parameters

This study has utilized the grid-search scheme to determine the optimal value of the regularization coefficient (C) and kernel parameter (σ) of the K-ELM classifier. The grid-search experiments train and test the classifier for different values of C and σ in the natural logarithmic scale from 1 to 10 at an interval of 1. Subsequently, the scheme selects the values of C and σ corresponding to the best mean accuracy of ten runs of 10-fold CV as the optimal values. We conducted the gird-search experiments separately on all the five FER datasets using HOG and BHOG features.

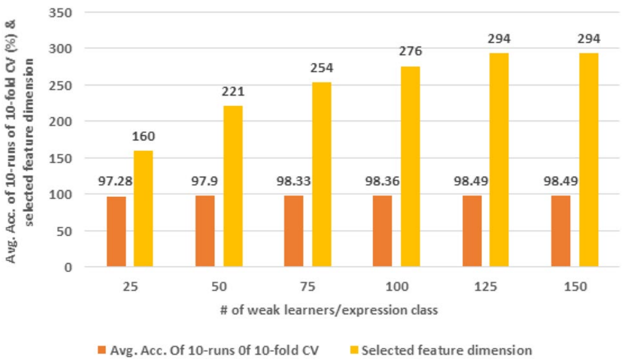

Fig. 10 Accuracy curves obtained by varying the number of weak learners or features per expression using AdaBoost FS algorithm and HOGv features extracted from the CK+ dataset

4.3 Evaluation results on the CK+ dataset

Table 1 reports the performance evaluation results of the proposed FER pipeline using the HOG and BHOG features on the CK+ dataset. The BHOG features with a much smaller feature dimension have achieved performance comparable to the original high-dimensional HOG features. Among the different variants, the HOGv variant using both directed and undirected orientation bins and the HOGs variant with only signed orientation bins achieved the best performance. Figure 12 shows the classification results of the proposed FER pipeline using AdaBoost selected HOGs features in terms of confusion matrix on the CK+ dataset. The pipeline correctly classified all the sample images belonging to the anger, disgust, fear, happiness, neutral, and sadness expression, while it wrongly classified one facial image from the surprise expression into neutral.

Upon examining the performance comparison results of Table 2, one can find that the proposed FER scheme using the BHOG features with K-ELM classifier achieved competitive accuracy compared to several state-of-the-art FER methods based on combinations of texture, shape, and hybrid texture & shape features [1, 10, 16, 20, 37, 45]. On the CK+ dataset, with 10-fold CV accuracy of 99.84%, the proposed FER pipeline has also performed better than the previous best recognition accuracy of 99.68% attained by the LBF-NN method [39]. Thus, using BHOG features, the proposed FER scheme significantly boosted the recognition accuracy of the proposed FER pipeline. Moreover, different variants of the BHOG features have also achieved performance better than existing deep-learning-based FER methods [46, 69–72].

4.4 Evaluation results on the JAFFE dataset

Table 3 reports the performance evaluation results of the proposed FER pipeline on the JAFFE dataset using features extracted by all three HOG and BHOG variants. On

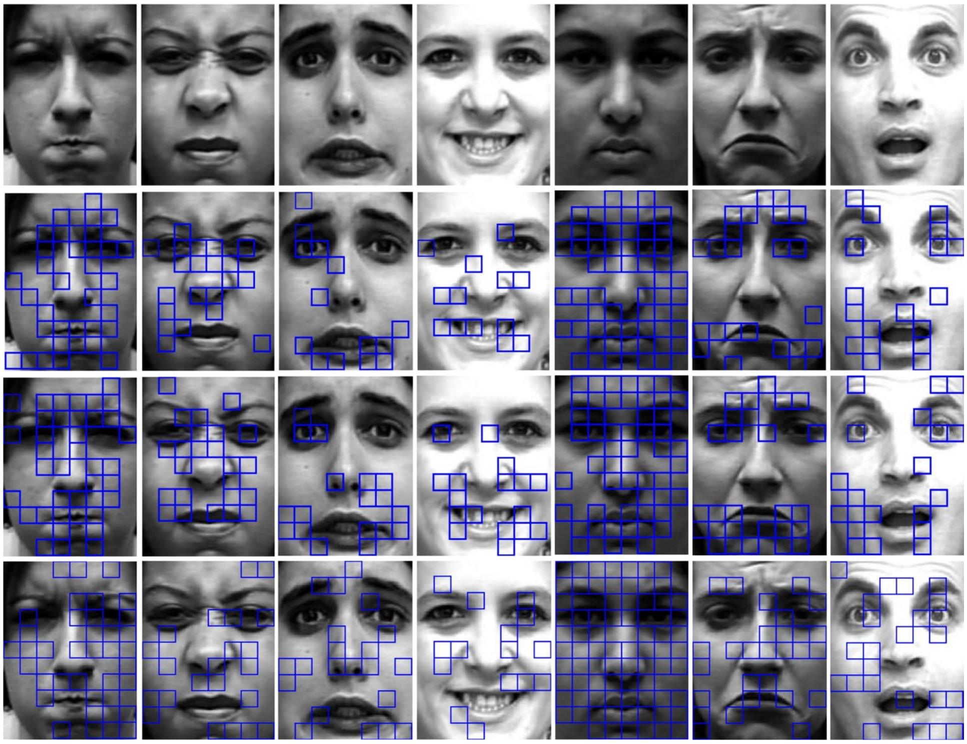

Fig. 11 AdaBoost selected facial cells on the CK+ sample facial images (top to bottom and left to right): Original facial expression images, HOGs selected cells, HOGu selected cells, and HOGv selected cells corresponding to anger, disgust, fear, happiness, neutral, sadness, and surprise expression

Table 1 Performance of proposed FER scheme on the CK+ dataset using 10-fold CV setting

Method	No FS		AdaBoost FS	
	Feature dimension	Accuracy (%)	Feature dimension	Accuracy (%)
HOGs	5880	99.76	276	**99.84**
HOGu	5880	99.76	303	99.76
HOGv	5896	**99.84**	294	**99.84**

Bold values represent the best result of the experiments

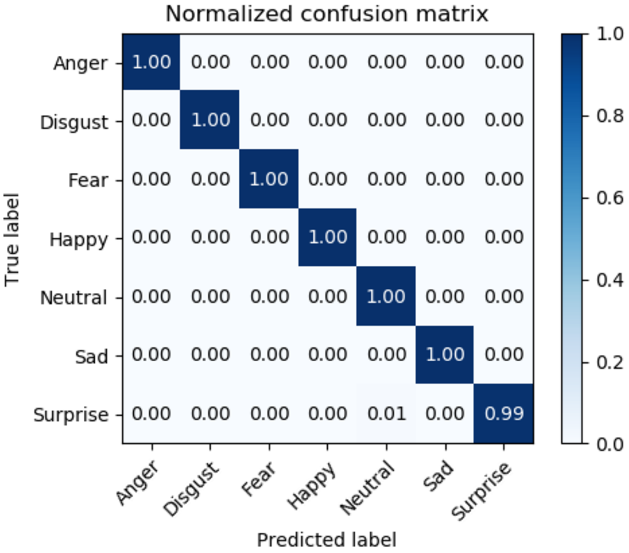

Fig. 12 Confusion matrix on the CK+ dataset using BHOGs descriptor

this dataset, the 168-dimensional BHOGs descriptor and 5896-dimensional HOGv descriptor achieved recognition accuracy of 98.63% and 98.61%, respectively. Thus, on the JAFFE dataset, too, the BHOG features with reduced dimensions achieved competing performance compared to the original high-dimensional HOG features. The results demonstrate the usefulness of the HOG features, AdaBoost FS algorithm, and K-ELM classifier in the FER task. Figure 13 shows the classification results of the proposed FER pipeline in terms of the confusion matrix on the JAFFE dataset using the BHOGs descriptor. The FER pipeline correctly classified all the facial images of disgust, fear, happiness, and neutral expression. Also, the proposed FER pipeline wrongly classifies one sample facial image from the anger, sadness, and surprise facial expressions.

The comparison results of Table 4 reveal that the proposed FER pipeline achieved competitive accuracy as compared to several related machine-learning-based FER methods [10, 16, 25, 32]. Using the 256-dimensional PCA-reduced IGLTP feature, the FER pipeline proposed by Holder and Tapamo [16] has achieved recognition accuracy of 81.70% on the JAFFE dataset. Also, the FER method using the combination of local binary features (LBF) and the neural network has reported achieving recognition accuracy of 98.10% [39]. In contrast, the proposed FER pipeline with

Table 2 Performance comparison with other state-of-the-art FER methods on the CK+ dataset

References	Technique	Feature dimensions	Accuracy (%)	Testing Protocol
2015 [1]	HOG + SVM	–	98.50	10-fold CV
2016 [10]	WLD + SVM	–	98.82	7-fold CV
2017 [37]	Multi-gradient EQP + SVM	–	99.36	10-fold CV
2017 [20]	LBP + HOG + SVM	11,636	98.30	10-fold CV
2017 [16]	GLTP + SVM	–	96.90	10-fold CV
2017 [16]	IGLTP + SVM	256	97.60	10-fold CV
2017 [45]	LTP + HOG + SVM	–	96.06	10-fold CV
2018 [46]	WMDNN	–	97.02	10-fold CV
2018 [69]	CNN Ensemble	–	95.36	10-fold CV
2018 [39]	LBF-NN	–	99.68	10-fold CV
2019 [32]	Gradient LPQ + SVM	–	97.05	10-fold CV
2019 [29]	ICLTP + K-NN + SRC	–	97.80	10-fold CV
2019 [70]	Deep learning technique	–	98.38	10-fold CV
2019 [71]	DAM-CNN	–	95.88	10-fold CV
2020 [72]	Deep Learning	–	97.38	10-fold CV
Proposed	HOGs + K-ELM	5880	99.76	10-fold CV
Proposed	HOGu + K-ELM	5880	99.76	10-fold CV
Proposed	HOGv + K-ELM	5896	99.84	10-fold CV
Proposed	BHOGs + K-ELM	276	**99.84**	10-fold CV
Proposed	BHOGu + K-ELM	303	99.76	10-fold CV
Proposed	BHOGv + K-ELM	294	99.84	10-fold CV

Bold value represent the best result of the experiment

Table 3 Performance of the proposed FER scheme on the JAFFE dataset using 10-fold CV setting

Method	No FS		AdaBoost FS	
	Feature dimension	Accuracy (%)	Feature dimension	Accuracy (%)
HOGs	5880	97.62	168	**98.63**
HOGu	5880	97.62	166	97.68
HOGv	5896	98.61	159	96.71

Bold value represent the best result of the experiment

just 168 BHOGs features has achieved superior recognition accuracy of 98.63% on the JAFFE dataset. Similar to the CK+ dataset, on the JAFFE dataset, too, the proposed FER pipeline achieved better accuracy than the state-of-the-art deep-learning-based FER techniques [29, 34, 46, 69, 72].

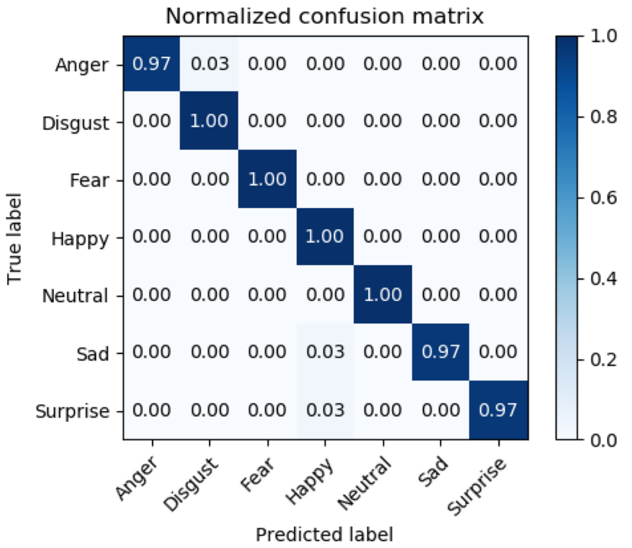

Fig. 13 Confusion matrix on the JAFFE dataset using BHOGs descriptor

4.5 Evaluation results on RaFD

Table 5 reports the performance evaluation results of the proposed FER scheme on the RaFD Category-1, RaFD Category-2, RaFD Category-3, and RaFD Category-4 datasets, using the 10-fold CV setting. These experiments have used original high-dimensional HOG features, and AdaBoost selected BHOG features. Upon examining the results of Table 5, one can find that on the RaFD Category-1 dataset, the 146-dimensional BHOGs feature performed well and achieved recognition accuracy of 100%. Similarly, on the RaFD Category-2 dataset, the BHOGv descriptor with a recognition accuracy of 99.58% delivered

Table 4 Performance comparison with other state-of-the-art FER methods on the JAFFE dataset

References	Technique	Feature dimensions	Accuracy (%)	Testing protocol
2016 [10]	WLD + SVM	–	97.00	7-fold CV
2017 [25]	K-ELBP + SVM	–	93.30	train-test split
2017 [16]	GLTP + SVM	256	74.40	10-fold CV
2017 [16]	IGLTP + SVM	256	81.70	10-fold CV
2018 [46]	WMDNN	–	92.21	10-fold CV
2018 [34]	LBI-CT	192	94.50	10-fold CV
2018 [69]	CNN Ensemble	–	96.57	10-fold CV
2018 [39]	LBF-NN	–	98.10	10-fold CV
2019 [32]	Gradient LPQ + SVM	–	92.19	10-fold CV
2019 [29]	ICLTP + K-NN + SRC	–	92.10	10-fold CV
2020 [72]	Deep Learning	–	97.18	10-fold CV
Proposed	HOGs + K-ELM	5880	97.62	10-fold CV
Proposed	HOGu + K-ELM	5880	97.62	10-fold CV
Proposed	HOGv + K-ELM	5896	98.61	10-fold CV
Proposed	BHOGs + K-ELM	168	**98.63**	10-fold CV
Proposed	BHOGu + K-ELM	166	97.68	10-fold CV
Proposed	BHOGv + K-ELM	159	96.71	10-fold CV

Bold value represent the best result of the experiment

Table 5 Performance of the proposed FER scheme on RaFD using 10-fold CV setting

Category	Method	No FS		AdaBoost FS	
		Feature dimension	Accuracy (%)	Feature dimension	Accuracy (%)
Category-1	HOGs	5880	99.37	146	**100.00**
	HOGu	5880	98.72	162	99.36
	HOGv	5896	99.57	159	99.37
Category-2	HOGs	5880	98.94	152	99.37
	HOGu	5880	98.54	173	98.75
	HOGv	5896	98.92	163	**99.58**
Category-3	HOGs	5880	97.78	242	98.69
	HOGu	5880	97.78	242	98.69
	HOGv	5896	98.51	239	**99.62**
Category-4	HOGs	5880	99.93	252	**99.93**
	HOGu	5880	99.93	294	**99.93**
	HOGv	5896	99.93	274	**99.93**

Bold values represent the best result of the experiments

the best performance among all descriptors. On the RaFD Category-3 dataset, the BHOGv features with a recognition accuracy of 99.62% achieved better performance than the original HOGv feature that has achieved a recognition accuracy of 98.51%. Finally, on the RaFD Category-4 dataset, all three variants of the HOG and BHOG descriptors achieved a recognition accuracy of 99.93%.

Table 6 Performance of proposed FER scheme using validation setting of [18, 19] on RaFD Category-5 dataset

Method	No FS		AdaBoost FS	
	Feature dimension	Accuracy (%)	Feature dimension	Accuracy (%)
HOGs	5880	97.92	246	99.24
HOGu	5880	96.78	333	98.30
HOGv	5896	98.67	309	98.30
HOGs+HOGu	–	–	579	**99.43**

Bold value represent the best result of the experiment

On the RaFD Category-5 dataset, following the standard procedure [18, 19], instead of the 10-fold CV, the proposed FER pipeline is trained and tested on the subject-independent train and test splits. Consequently, we divided the dataset into a 2:1 ratio. Out of the 67 subjects, the training set includes the facial images of 45 subjects, and the test set contains the rest of the 22 subjects' facial images. We ensured that the same subject did not fall in both training and test sets during the distribution. Table 6 reports the evaluation results of the different HOG variants with and without feature selection on the RaFD Category-5 dataset. On this dataset, the BHOG descriptors performed better than the HOG descriptors. Also, in contrast to the standalone BHOGs and BHOGu features, their fused variant with a recognition accuracy of 99.43% attained the best performance. Figure 14 shows the classification result of the proposed FER pipeline

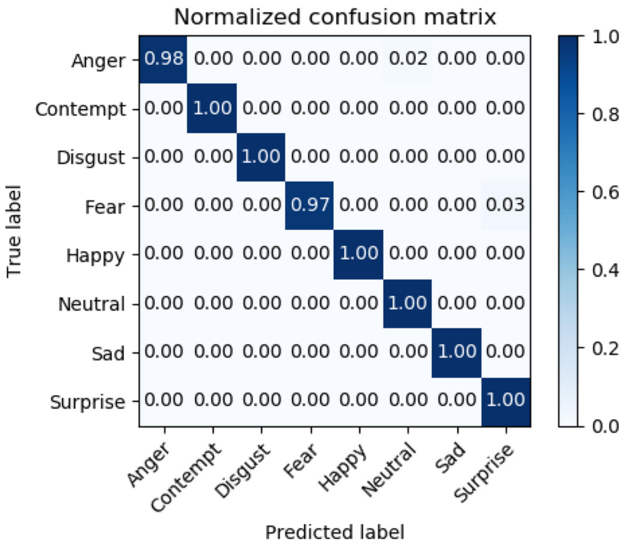

Fig. 14 Confusion matrix on RaFD Category-5 dataset using fusion of BHOGs and BHOGu descriptor

using the fusion of the BHOGs and BHOGu features in terms of the confusion matrix on the test set of the RaFD Category-5 dataset. The pipeline correctly classified all the facial images of contempt, disgust, happiness, neutral, sadness, and surprise. However, the pipeline wrongly classified one facial image from the anger class as neutral and two facial images from the fearful class as a surprise.

Based on the performance comparison results of Table 7, one can notice that on the RaFD Category-1, RaFD Category-3, and RaFD Category-4 datasets, the proposed FER scheme using the 10-fold CV setting achieved superior recognition accuracy compared to the available state-of-the-art FER methods based on the traditional machine-learning [1] and deep-learning [70]. Moreover, the performance comparison results of Table 8 show that on the RaFD Category-5 dataset, adopting the subject-independent evaluation protocol, the proposed FER pipeline using BHOG features attained superior classification performance. Also, on the RaFD Category-5 dataset, the FER framework proposed by Ghosh et al. [19] has delivered a classification score of 99.25% using the combination of Gabor feature extractor, Late Hill Climbing-based Memetic Algorithm (LHCMA) based feature selection algorithm, and the Sequential Minimal Optimization (SMO) classifier. The FER pipeline proposed by Saha et al. [18], on the other hand, using Gabor feature extractor, Supervised Filter Harmony Search Algorithm (SFHSA) based feature selection algorithm, and SMO classifier has attained a recognition accuracy of 97.79%. Nevertheless, using the fusion of BHOGs and BHOGu features, the proposed FER pipeline with a recognition accuracy of 99.43% has surpassed the accuracy reported by both Ghosh et al. [19] and Saha et al. [18]. These results confirm the effectiveness of the proposed FER pipeline using the HOG

Table 7 Performance comparison with other state-of-the-art FER methods on RaFD using 10-fold CV setting

References	Technique	Feature dimension	Accuracy (%)	RaFD category
2015 [1]	HOG + SVM	–	94.90	Category-1
2015 [1]	HOG + SVM	–	92.90	Category-3
2019 [70]	Deep learning	–	99.17	Category-4
Proposed	BHOGs+K-ELM	146	**100.00**	Category-1
Proposed	BHOGv+K-ELM	163	**99.58**	Category-2
Proposed	BHOGv+K-ELM	239	**99.62**	Category-3
Proposed	BHOGv+K-ELM	274	**99.93**	Category-4

Bold values represent the best result of the experiments

Table 8 Performance comparison with other state-of-the-art FER methods on RaFD Category-5 dataset using the validation setting of [18, 19]

References	Technique	Feature dimension	Accuracy (%)
2019 [19]	Gabor + SA + SMO	1333	95.90
2019 [19]	Gabor + GA + SMO	1613	98.32
2019 [19]	Gabor + MA + SMO	894	98.88
2019 [19]	Gabor + ME-BPSO + SMO	1300	98.13
2019 [19]	Gabor + WOA-CM + SMO	1186	97.01
2019 [19]	Gabor + LHCMA + SMO	1271	99.25
2019 [18]	Gabor + SFHSA + SMO	462	97.79
2019 [19]	HOG + SA + SMO	816	92.66
2019 [19]	HOG + GA + SMO	480	97.01
2019 [19]	HOG + MA + SMO	390	97.01
2019 [19]	HOG + ME-BPSO + SMO	1039	95.15
2019 [19]	HOG + WOA-CM + SMO	1041	94.96
2019 [19]	HOG + LHCMA + SMO	800	97.57
2019 [18]	HOG + SFHSA + SMO	544	96.32
Proposed (Ours)	BHOGs + K-ELM	246	99.24
Proposed (Ours)	BHOGu + K-ELM	333	98.30
Proposed (Ours)	BHOGv + K-ELM	309	98.30
Proposed (Ours)	BHOGs + BHOGu + K-ELM	**579**	**99.43**

Bold values represent the best result of the experiments

Table 9 Age-independent analysis results on the TFE dataset

Method	No FS		AdaBoost FS	
	Feature dimension	Accuracy (%)	Feature dimension	Accuracy (%)
HOGs	5880	86.10	352	84.41
HOGu	5880	86.10	376	84.07
HOGv	5896	87.12	357	85.42
HOGv + HOGs	–	–	709	**87.46**

Bold value represent the best result of the experiment

Table 10 Age-dependent analysis results on the TFE dataset

Scenario	Method	No FS		AdaBoost FS	
		Feature dimension	Accuracy (%)	Feature dimension	Accuracy (%)
Scenario-1	HOGs	5880	83.47	316	**84.86**
	HOGu	5880	79.68	334	78.69
	HOGv	5896	81.27	298	77.09
Scenario-2	HOGs	5880	**76.33**	260	74.20
	HOGu	5880	69.68	281	69.95
	HOGv	5896	75.80	270	73.94

Bold values represent the best result of the experiments

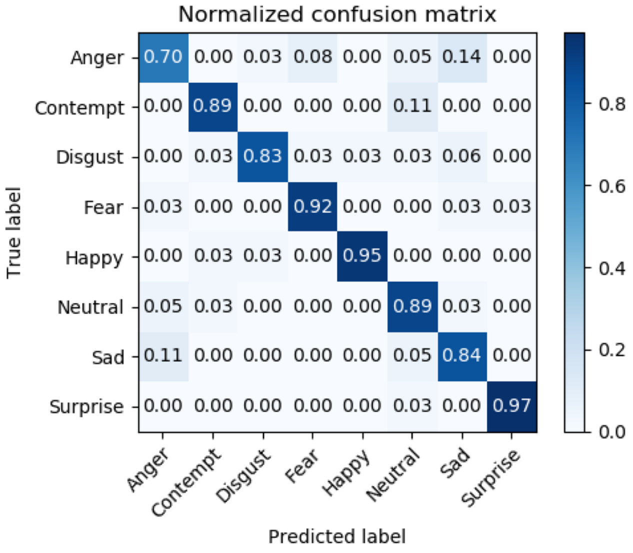

Fig. 15 Confusion matrix result of age-independent expression analysis on the TFE dataset using fusion of BHOGs and BHOGv descriptor

descriptor, AdaBoost FS algorithm, and K-ELM classifier.

4.6 Evaluation results on the TFE dataset

Table 9 shows the results of the age-independent expression analysis of the proposed FER pipeline on the TFE dataset. On the age-independent test set of the TFE dataset, among different variants of HOG and BHOG descriptors, the fusion of the BHOGv and BHOGs with a test accuracy of 87.46% achieved the best performance. Figure 15 shows the confusion matrix on the test set of the TFE dataset. Among the eight prototypical facial expressions, happiness (95% accuracy), fear (92% accuracy), and surprise (97% accuracy) were the most accurately classified expressions. The fused BHOG features also achieved satisfactory performance in classifying facial images belonging to neutral (89%), sad (84%), disgust (83%), and contempt (89%). Out of the eight classes, anger, with a classification accuracy of 70%, was the least correctly recognized expression. A major fraction (14%) of the misclassified images from the anger expression got classified as sadness.

The age-dependent analysis results reported in Table 10 show that the proposed FER pipeline trained on the BHOGs features extracted from the facial images of older adults achieved 84.86% recognition accuracy on the facial test images of young adults. Meanwhile, the pipeline trained on HOGs features extracted from the facial images of young adults attained recognition of 76.33% on the facial test images of older adults. Thus, the age-dependent analysis results reveal that the facial expressions of older adults are more prominent than the young adults.

4.7 Evaluation results on RAF-DB

RAF-DB has widely been used to validate the performance of the deep learning and traditional machine learning FER techniques [17, 31, 73–78].

Table 11 reports the performance evaluation results of the proposed FER pipeline on RAF-DB using features extracted by all three HOG (HOGs, HOGu, and HOGv) and BHOG (BHOGs, BHOGu, BHOGv) variants. On the validation set of RAF-DB, 5880-dimensional HOGs descriptor with a recognition accuracy of 80.51% attained the highest accuracy. Besides, the BHOGs descriptor with just 786-dimension attained the second-highest recognition accuracy of 80.15% on the validation set of RAF-DB. Thus, on the FER in-the-wild RAF-DB dataset, too, the BHOG descriptors with reduced dimensions achieved competing performance compared to the original high-dimensional HOG descriptors. Figure 16 shows the confusion matrix on the validation set of RAF-DB using HOGs descriptor. The proposed FER pipeline correctly classified 95%, 84%, 73%, 77%, and 65% of the facial image samples belonging to happiness, neutral, sadness, surprise, and anger. The proposed pipeline with a classification accuracy of 29% and 38% on the disgust and

Table 11 Performance of the proposed FER scheme on RAF-DB

Method	No FS		AdaBoost FS	
	Feature dimension	Accuracy (%)	Feature dimension	Accuracy (%)
HOGs	5880	**80.51**	786	80.15
HOGu	5880	77.74	793	77.48
HOGv	5896	79.24	767	78.75

Bold value represent the best result of the experiment

fearful expression, respectively, performed poorly on these two facial expressions.

Table 12 reports the performance comparison results of the proposed FER pipeline using BHOG features with the other state-of-the-art FER methods on RAF-DB. It is not surprising that on the large-scale FER in the wild dataset like RAF-DB, the existing deep learning-based techniques [17, 74–79] performed much better than the traditional machine learning methods [31, 73, 76]. The results are consistent with the fact that the classification accuracy of the deep learning classification algorithms is directly proportional to the size of the dataset, while the classification accuracy of the image classification pipeline based on traditional machine learning gets saturated after training on a certain size of the dataset and does not improve if we further increase the size of the dataset. This might be the reason why the vast majority of works related to FER based on traditional machine learning use the FER in the lab datasets such as CK+ [64], JAFFE [65], RaFD [66], KDEF [80], and TFE [67] to validate the performance of the FER pipeline. Only a small fraction of the works on FER using traditional machine learning [31, 73, 76] have used large-scale FER in the wild dataset like RAF-DB [73] to validate their performance.

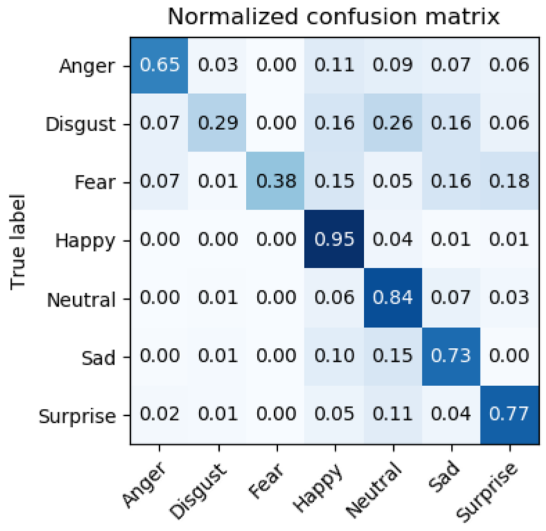

Fig. 16 Confusion matrix on RAF-DB using HOGs descriptor

Table 12 Performance comparison with other state-of-the-art FER methods on RAF-DB

References	Technique	Feature dimensions	Accuracy (%)
2018 [68]	LBP + SVM	–	72.71
2018 [68]	Gabor + SVM	–	74.35
2018 [68]	HOG + SVM	–	77.28
2021 [31]	DLTP + K-ELM	45056	78.75
2021 [31]	uDLTP + K-ELM	10384	78.46
2018 [68]	DLP-CNN	–	84.13
2021 [74]	DAN	–	89.70
2021 [75]	FER-VT	–	88.26
2021 [17]	EmNet	–	87.16
2022 [79]	ARM	–	90.42
2022 [76]	VGG	–	85.72
2022 [76]	SENet	–	86.31
2022 [76]	DenseNet	–	84.09
2022 [76]	Xception	–	82.65
2022 [76]	LBP + SVM	–	75.08
2022 [78]	CNN	–	87.35
2022 [77]	DICNN	–	86.07
Proposed	HOGs + K-ELM	5880	80.51
Proposed	HOGu + K-ELM	5880	77.74
Proposed	HOGv + K-ELM	5896	79.24
Proposed	BHOGs + K-ELM	786	80.15
Proposed	BHOGu + K-ELM	793	77.48
Proposed	BHOGv + K-ELM	767	78.75

On the validation set of RAF-DB, the traditional machine learning-based FER pipeline proposed by Li and Deng [68] has attained classification accuracy of 72.71%, 74.35%, and 77.28% using LBP, HOG, and Gabor descriptor, respectively. Also, the FER pipeline proposed by Greco et al. [76] has attained a recognition accuracy of 75.08% using the combination of LBP descriptor and SVM classifier. Besides, on the validation set of RAF-DB, the traditional machine learning-based FER pipeline introduced by Saurav et al. [31] has achieved recognition accuracy of 78.75% and 78.46% using the DLTP descriptor and the uniform DLTP descriptor, respectively. Among the existing deep learning-based FER techniques, on the validation set of RAF-DB, the Amending Representation Module (ARM) introduced by Shi et al. [79] has attained the highest recognition accuracy of 90.42%. Besides, evaluated on the facial images in the the validation set of RAF-DB, Distract your Attention Network (DAN) [74], deep locality-preserving convolutional neural network (DLP-CNN) [68], FER-VT [75], EmNet [17], and DICNN [77] has reported achieving classification accuracy of 89.70%, 84.13%, 88.26%, 87.16%, and

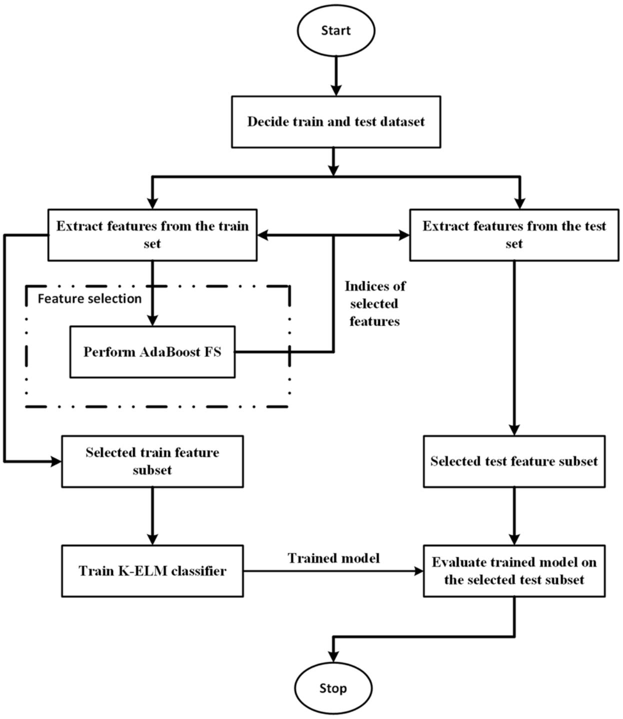
Start
Decide train and test dataset
Extract features from the train set
Extract features from the test set
Feature selection
Perform AdaBoost FS
Indices of selected features
Selected train feature subset
Selected test feature subset
Train K-ELM classifier
Trained model
Evaluate trained model on the selected test subset
Stop

Fig. 17 Flowchart representing the cross-dataset evaluation scheme

86.07%, respectively. Trained and tested on RAF-DB, the proposed FER pipeline using HOGs + K-ELM and BHOGs + K-ELM has attained a competitive recognition accuracy of 80.51% and 80.15%, respectively. Thus, on RAF-DB, the performance of the proposed FER pipeline has surpassed the performance reported by the exiting FER methods based on traditional machine learning. Also, the proposed framework is computationally more efficient than the exiting FER methods based on deep learning. In summary, as compared to the existing FER methods, the proposed FER pipeline is compute-efficient and robust and thus suitable for real-time recognition of facial expressions running on a resource-constrained embedded platform.

4.8 Cross-dataset performance evaluation

Apart from the 10-fold CV and subject-independent train-test evaluation procedure, this study also conducted the cross-dataset evaluation to assess the generalization of the FER pipeline introduced in this work. The cross-dataset evaluation, as shown in Fig. 17 uses one FER dataset as the train set and the other as the test set. In cross-dataset testing, once the train and test FER datasets are decided, in the subsequent step, features are extracted using all the three variants of the HOG descriptor from both the train and test datasets. Subsequently, the training set features are fed to the AdaBoost FS algorithm to discover optimal features from the original high-dimensional features. Once trained, based on the indices of the AdaBoost selected features, the FS scheme extracts BHOG features from the pre-computed train and test features. Finally, the cross-dataset scheme train and test the K-ELM classifier on the train and test BHOG features using the optimal values of the kernel parameter and regularization factor determined using the grid-search scheme.

Table 13 Cross-dataset performance evaluation results of the proposed FER scheme

Train	Test	Method	Expressions	Accuracy (%)
CK+	RaFD	BHOGs (HOGs)	7	90.41 (**90.83**)
		BHOGu (HOGu)	7	84.43 (89.13)
		BHOGv (HOGv)	7	89.34 (90.41)
		Deep Learning [70]	7	86.80
CK+	JAFFE	BHOGs (HOGs)	7	33.80 (36.15)
		BHOGu (HOGu)	7	41.31 (43.66)
		BHOGv (HOGv)	7	43.19 (35.31)
		Deep Learning [72]	6	39.01
		DAM-CNN [71]	6	**43.38**
		LBP-SVM [15]	7	41.30
		CCEC [81]	6	42.30
RaFD	CK+	BHOGs (HOGs)	7	86.89 (**89.48**)
		BHOGu (HOGu)	7	84.39 (89.24)
		BHOGv (HOGv)	7	86.08 (86.73)
		Deep Learning [70]	7	75.13
RaFD	JAFFE	BHOGs (HOGs)	7	42.72 (40.38)
		BHOGu (HOGu)	7	**49.77** (41.31)
		BHOGv (HOGv)	7	42.25 (41.31)
JAFFE	CK+	BHOGs (HOGs)	7	41.75 (51.05)
		BHOGu (HOGu)	7	49.27 (**57.04**)
		BHOGv (HOGv)	7	44.66 (55.50)
		DAM-CNN [71]	6	49.10
		CCEC [81]	6	48.20
		DCMA-CNNs [82]	6	46.28
JAFFE	RaFD	BHOGs (HOGs)	7	49.47 (**55.86**)
		BHOGu (HOGu)	7	51.60 (55.22)
		BHOGv (HOGv)	7	50.53 (53.30)

Bold values represent the best result of the experiments

Table 13 summarizes the results of cross-dataset experiments. The proposed FER scheme using the HOG and BHOG descriptor achieved competitive cross-dataset test accuracy. Utilizing the CK+ dataset as the train set and the RaFD Category-2 dataset as the test set, the proposed FER scheme using HOGs and BHOGs features achieved test accuracy better than the deep learning-based FER technique [70]. The deep-learning-based FER technique introduced by Sun et al. [70] has achieved a test accuracy of 86.80%, whereas the proposed FER method with a test accuracy of 90.83% has registered a boost of 4.03%. Similarly, using the CK+ dataset as a train set and the JAFFE dataset as the

Table 14 Computation time comparison in milliseconds (ms) (DR: Dimensionality reduction and FS: Feature selection)

Method	CPU	Feature extraction	Classification without DR/FS	Classification with DR/FS	Total
LBP [48]	Octa-core 3.50 GHz	28.9	76.7	34.5	111.20
LDP [48]	Octa-core 3.50 GHz	106.9	81.4	37.2	225.50
LTP [48]	Octa-core 3.50 GHz	51.4	146.5	36.5	234.40
LDN [48]	Octa-core 3.50 GHz	132.9	73.2	33.8	239.90
LNEP [48]	Octa-core 3.50 GHz	22.9	76.7	34.6	134.40
LBP [27]	Intel i5 3.2 GHz	–	–	–	295.50
LBF-NN [39]	Intel i7-7500U 2.70 GHz	–	–	–	**1.00**
LBP [15]	–	–	–	–	30.00
HOGu (Ours)	Intel i9-8950HK 2.90 GHz	1.1	1.05	0.32	1.42
HOGs (Ours)	Intel i9-8950HK 2.90 GHz	1.1	1.18	0.33	1.43
HOGv (Ours)	Intel i9-8950HK 2.90 GHz	1.1	1.11	0.30	1.40

Bold value represent the best result of the experiment

test set, the BHOGv features achieved the highest recognition accuracy of 43.19%. However, on this train and test combination of the FER datasets, the deep learning-based FER technique of Xie et al. [71] has registered the best test accuracy of 43.38%. The low cross-dataset accuracy of the proposed method may be due to cultural bias in the train and test combination of the FER datasets.

Besides, on the train and test combinations of RaFD Category-2 and CK+ dataset, the proposed FER pipeline with a recognition accuracy of 89.48% has attained a boost in recognition accuracy of 14.35%, over the previous best test accuracy of 75.13% reported by the deep-learning-based FER method [70]. Using the JAFFE dataset as a train set and CK+ dataset as the test set, the HOGu features attained a test accuracy of 57.04%, better than several existing FER methods [71, 81, 82]. Finally, the train and test combination of JAFFE and RaFD using HOGs feature attained the best cross-dataset test accuracy of 55.86%. In summary, in addition to 10-fold CV accuracy, the proposed FER pipeline has also achieved better cross-dataset accuracy on different combinations of the FER datasets. Thus, the proposed pipeline shows better generalization performance than the state-of-the-art FER methods.

5 Computational performance analysis

Table 14 demonstrates the comparison of computation time for feature extraction using the HOG/BHOG descriptors and classification by the K-ELM classifier. We conducted the feature selection experiment offline using the AdaBoost FS algorithm and used the selected features to calculate the classification time.

The K-ELM classifier, as expected, takes less time to classify the AdaBoost-selected HOG features than the original high-dimensional HOG features. It attains a four times boost in the execution speed using BHOG features. Besides, feature extraction by different HOG variants is also more efficient than the previously reported feature extraction schemes using the LBP, LDP, LTP, LDN, and LNEP descriptors [15, 27, 48]. Among the reported methods, the FER scheme using LBF-NN with total feature extraction and classification time of 1.00 ms is computationally the most efficient [39]. Nevertheless, using BHOG features and the K-ELM classifier, the proposed FER pipeline takes 1.40 ms to classify an input facial image. Thus, the proposed FER pipeline using boosted features is computationally efficient and thus suitable for real-time applications.

6 Conclusions

This study presented a reliable and computationally efficient method for FER utilizing the Boosted Histogram of oriented Gradient (BHOG) features. The proposed FER pipeline has utilized a face alignment & registration unit to get a well-aligned and registered facial image of standard size. Subsequently, it extracted facial features using different HOG descriptors from the registered facial images. Feature selection using the AdaBoost FS algorithm helped the pipeline eliminate irrelevant features from the high-dimensional HOG features. Finally, using the K-ELM classifier, the proposed FER scheme classified the features into facial expressions. We evaluated the performance of the proposed FER pipeline on five FER datasets, namely the CK+, JAFFE, RaFD, TFE, and RAF-DB, using three different testing procedures, namely the 10-fold CV, subject-independent train-test split, and cross-dataset evaluation. Performance comparison results with existing FER techniques exhibited the effectiveness of the proposed FER pipeline. The BHOG features helped the pipeline achieve competitive recognition accuracy with a multi-fold improvement in the overall computational time. Thus, the proposed FER pipeline is a suitable candidate for the real-time classification of facial expressions. The extended version of the work will deal with the fusion of BHOG features with other facial textures and geometric features. Further, the method can be tested on the other FER in the wild datasets to evaluate their robustness in complex real-world conditions.

Acknowledgements The authors would like to thank the director, CSIR-CEERI, Pilani for supporting and encouraging research activities at CSIR-CEERI, Pilani. Constant motivation by the group head of the cognitive computing group at CSIR-CEERI is also acknowledged.

Funding No external funding.

Data availability All dataset used in the study are freely available.

Code availability Custom Code.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Carcagnì P, Del Coco M, Leo M, Distante C (2015) Facial expression recognition and histograms of oriented gradients: a comprehensive study. Springerplus 4(1):645
2. Jian BL, Chen CL, Chu WL, Huang MW (2017) The facial expression of schizophrenic patients applied with infrared thermal facial image sequence. BMC Psychiatry 17(1):1–7
3. Uddin MZ, Hassan MM, Almogren A, Alamri A, Alrubaian M, Fortino G (2017) Facial expression recognition utilizing local direction-based robust features and deep belief network. IEEE Access 5:4525–4536
4. Zhao J, Mao X, Chen L (2019) Speech emotion recognition using deep 1d & 2d cnn lstm networks. Biomed Signal Process Control 47:312–323
5. Mehdizadehfar V, Ghassemi F, Fallah A, Pouretemad H (2020) Eeg study of facial emotion recognition in the fathers of autistic children. Biomed Signal Process Control 56:101721
6. Oh S, Lee JY, Kim DK (2020) The design of cnn architectures for optimal six basic emotion classification using multiple physiological signals. Sensors 20(3):866
7. Avots E, Sapiński T, Bachmann M, Kamińska D (2019) Audiovisual emotion recognition in wild. Mach Vis Appl 30(5):975–985
8. Huang Y, Yang J, Liu S, Pan J (2019) Combining facial expressions and electroencephalography to enhance emotion recognition. Future Internet 11(5):105
9. Huang X, Kortelainen J, Zhao G, Li X, Moilanen A, Seppänen T, Pietikäinen M (2016) Multi-modal emotion analysis from facial expressions and electroencephalogram. Comput Vis Image Underst 147:114–124
10. Alhussein M (2016) Automatic facial emotion recognition using weber local descriptor for e-healthcare system. Clust Comput 19(1):99–108
11. Jeong M, Ko BC (2018) Driver's facial expression recognition in real-time for safe driving. Sensors 18(12):4270
12. Sini J, Marceddu AC, Violante M (2020) Automatic emotion recognition for the calibration of autonomous driving functions. Electronics 9(3):518
13. Li THS, Kuo PH, Tsai TN, Luan PC (2019) Cnn and lstm based facial expression analysis model for a humanoid robot. IEEE Access 7:93998–94011
14. Yolcu G, Oztel I, Kazan S, Oz C, Palaniappan K, Lever TE, Bunyak F (2019) Facial expression recognition for monitoring neurological disorders based on convolutional neural network. Multimed Tools Appl 78(22):31581–31603
15. Shan C, Gong S, McOwan PW (2009) Facial expression recognition based on local binary patterns: A comprehensive study. Image Vis Comput 27(6):803–816
16. Holder RP (2017) Tapamo JR (2017) Improved gradient local ternary patterns for facial expression recognition. EURASIP J Image Video Process 1:42
17. Saurav S, Saini R, Singh S (2021) Emnet: a deep integrated convolutional neural network for facial emotion recognition in the wild. Appl Intell 51(8):5543–5570
18. Saha S, Ghosh M, Ghosh S, Sen S, Singh PK, Geem ZW, Sarkar R (2020) Feature selection for facial emotion recognition using cosine similarity-based harmony search algorithm. Appl Sci 10(8):2816
19. Ghosh M, Kundu T, Ghosh D, Sarkar R (2019) Feature selection for facial emotion recognition using late hill-climbing based memetic algorithm. Multimed Tools Appl 78(18):25753–25779
20. Liu Y, Li Y, Ma X, Song R (2017) Facial expression recognition with fusion features extracted from salient facial areas. Sensors 17(4):712
21. Saurav S, Singh S, Saini R, Yadav M (2020) Facial expression recognition using improved adaptive local ternary pattern. In: Proceedings of 3rd international conference on computer vision and image processing, Springer, pp 39–52
22. Revina IM, Emmanuel WS (2019) Face expression recognition with the optimization based multi-svnn classifier and the modified ldp features. J Vis Commun Image Represent 62:43–55
23. Rivera AR, Castillo JR, Chae OO (2012) Local directional number pattern for face analysis: face and expression recognition. IEEE Trans Image Process 22(5):1740–1752
24. Feng X, Pietikainen M, Hadid A (2005) Facial expression recognition with local binary patterns and linear programming. Pattern Recognit Image Anal C/C Raspoznavaniye Obrazov I Analiz Izobrazhenii 15(2):546
25. Guo M, Hou X, Ma Y, Wu X (2017) Facial expression recognition using elbp based on covariance matrix transform in klt. Multimed Tools Appl 76(2):2995–3010
26. Wang W, Chang F, Liu Y, Wu X (2017) Expression recognition method based on evidence theory and local texture. Multimed Tools Appl 76(5):7365–7379
27. Happy S, Routray A (2014) Automatic facial expression recognition using features of salient facial patches. IEEE Trans Affect Comput 6(1):1–12
28. Ahmed F, Hossain E (2013) Automated facial expression recognition using gradient-based ternary texture patterns. Chin J Eng 2013:831787
29. Luo Y, Xy Liu, Zhang Y, Xf Chen, Chen Z (2019) Facial expression recognition based on improved completed local ternary patterns. Optoelectron Lett 15(3):224–230
30. Iqbal MTB, Abdullah-Al-Wadud M, Ryu B, Makhmudkhujaev F, Chae O (2018) Facial expression recognition with neighborhood-aware edge directional pattern (nedp). IEEE Trans Affect Comput 11(1):125–137
31. Saurav S, Saini R, Singh S (2021) Facial expression recognition using dynamic local ternary patterns with kernel extreme learning machine classifier. IEEE Access 9:120844–120868

32. Kherchaoui S, Houacine A (2019) Facial expression identification using gradient local phase. Multimed Tools Appl 78(12):16843–16859
33. Ryu B, Rivera AR, Kim J, Chae O (2017) Local directional ternary pattern for facial expression recognition. IEEE Trans Image Process 26(12):6006–6018
34. Khan SA, Hussain A, Usman M (2018) Reliable facial expression recognition for multi-scale images using weber local binary image based cosine transform features. Multimed Tools Appl 77(1):1133–1165
35. Mahmood A, Hussain S, Iqbal K, Elkilani WS (2019) Recognition of facial expressions under varying conditions using dual-feature fusion. Math Prob Eng
36. Siddiqi MH, Ali R, Idris M, Khan AM, Kim ES, Whang MC, Lee S (2016) Human facial expression recognition using curvelet feature extraction and normalized mutual information feature selection. Multimed Tools Appl 75(2):935–959
37. Al-Sumaidaee SA, Abdullah MA, Al-Nima RRO, Dlay SS, Chambers JA (2017) Multi-gradient features and elongated quinary pattern encoding for image-based facial expression recognition. Pattern Recogn 71:249–263
38. Alphonse AS, Starvin M (2019) A novel maximum and minimum response-based gabor (mmrg) feature extraction method for facial expression recognition. Multimed Tools Appl 78(16):23369–23397
39. Gogić I, Manhart M, Pandžić IS, Ahlberg J (2018) Fast facial expression recognition using local binary features and shallow neural networks. Visual Comput 36(1):97–112
40. Revina IM, Emmanuel WS (2019) Mdtp: a novel multi-directional triangles pattern for face expression recognition. Multimed Tools Appl 78(18):26223–26238
41. Mlakar U, Potočnik B (2015) Automated facial expression recognition based on histograms of oriented gradient feature vector differences. SIViP 9(1):245–253
42. Nazir M, Jan Z, Sajjad M (2018) Facial expression recognition using histogram of oriented gradients based transformed features. Clust Comput 21(1):539–548
43. Nigam S, Singh R, Misra A (2018) Efficient facial expression recognition using histogram of oriented gradients in wavelet domain. Multimed Tools Appl 77(21):28725–28747
44. Zhang X, Mahoor MH, Mavadati SM (2015) Facial expression recognition using l_p-norm mkl multiclass-svm. Mach Vis Appl 26(4):467–483
45. Lekdioui K, Messoussi R, Ruichek Y, Chaabi Y, Touahni R (2017) Facial decomposition for expression recognition using texture/shape descriptors and svm classifier. Signal Process Image Commun 58:300–312
46. Yang B, Cao J, Ni R, Zhang Y (2017) Facial expression recognition using weighted mixture deep neural network based on double-channel facial images. IEEE Access 6:4630–4640
47. Ghimire D, Jeong S, Lee J, Park SH (2017) Facial expression recognition based on local region specific features and support vector machines. Multimed Tools Appl 76(6):7803–7821
48. Shanthi P, Nickolas S (2020) An efficient automatic facial expression recognition using local neighborhood feature fusion. Multimed Tools Appl 80(7):10187–10212
49. Lajevardi SM, Hussain ZM (2010) Novel higher-order local autocorrelation-like feature extraction methodology for facial expression recognition. IET Image Proc 4(2):114–119
50. Lajevardi SM, Hussain ZM (2012) Automatic facial expression recognition: feature extraction and selection. SIViP 6(1):159–169
51. Viola P, Jones MJ (2004) Robust real-time face detection. Int J Comput Vision 57(2):137–154
52. Xiong X, De la Torre F (2013) Supervised descent method and its applications to face alignment. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 532–539
53. Köstinger M (2013) Efficient metric learning for real-world face recognition
54. Felzenszwalb PF, Girshick RB, McAllester D, Ramanan D (2009) Object detection with discriminatively trained part-based models. IEEE Trans Pattern Anal Mach Intell 32(9):1627–1645
55. Hahnle M, Saxen F, Hisung M, Brunsmann U, Doll K (2013) Fpga-based real-time pedestrian detection on high-resolution images. In: 2013 IEEE conference on computer vision and pattern recognition workshops, IEEE, pp 629–635
56. Chen PY, Huang CC, Lien CY, Tsai YH (2013) An efficient hardware implementation of hog feature extraction for human detection. IEEE Trans Intell Transp Syst 15(2):656–662
57. Huang Z, Yu Y, Gu J, Liu H (2016) An efficient method for traffic sign recognition based on extreme learning machine. IEEE Trans Cybern 47(4):920–933
58. Chandrashekar G, Sahin F (2014) A survey on feature selection methods. Comput Electrical Eng 40(1):16–28
59. Gu Q, Li Z, Han J (2012) Generalized fisher score for feature selection. arXiv preprint arXiv:1202.3725
60. Schapire RE, Freund Y (2013) Boosting: Foundations and algorithms. Kybernetes
61. Huang GB, Zhou H, Ding X, Zhang R (2011) Extreme learning machine for regression and multiclass classification. IEEE Trans Syst Man Cybern Part B (Cybernetics) 42(2):513–529
62. Zeng Y, Xu X, Shen D, Fang Y, Xiao Z (2016) Traffic sign recognition using kernel extreme learning machines with deep perceptual features. IEEE Trans Intell Transp Syst 18(6):1647–1653
63. Vedaldi A, Fulkerson B (2010) Vlfeat: An open and portable library of computer vision algorithms. In: Proceedings of the 18th ACM international conference on Multimedia, pp 1469–1472
64. Lucey P, Cohn JF, Kanade T, Saragih J, Ambadar Z, Matthews I (2010) 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, IEEE, pp 94–101
65. Lyons MJ, Akamatsu S, Kamachi M, Gyoba J, Budynek J (1998) The japanese female facial expression (jaffe) database. In: Proceedings of third international conference on automatic face and gesture recognition, pp 14–16
66. Langner O, Dotsch R, Bijlstra G, Wigboldus DH, Hawk ST, Van Knippenberg A (2010) Presentation and validation of the radboud faces database. Cogn Emot 24(8):1377–1388
67. Yang T, Yang Z, Xu G, Gao D, Zhang Z, Wang H, Liu S, Han L, Zhu Z, Tian Y et al (2020) Tsinghua facial expression database-a database of facial expressions in chinese young and older women and men: Development and validation. PLoS ONE 15(4):e0231304
68. Li S, Deng W (2018) Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition. IEEE Trans Image Process 28(1):356–370
69. Sun A, Li Y, Huang YM, Li Q, Lu G (2018) Facial expression recognition using optimized active regions. HCIS 8(1):33
70. Sun N, Li Q, Huan R, Liu J, Han G (2019) Deep spatial-temporal feature fusion for facial expression recognition in static images. Pattern Recogn Lett 119:49–61
71. Xie S, Hu H, Wu Y (2019) Deep multi-path convolutional neural network joint with salient region attention for facial expression recognition. Pattern Recogn 92:177–191
72. Li K, Jin Y, Akram MW, Han R, Chen J (2020) Facial expression recognition with convolutional neural networks via a new face cropping and rotation strategy. Vis Comput 36(2):391–404

73. Li M, Xu H, Huang X, Song Z, Liu X, Li X (2018) Facial expression recognition with identity and emotion joint learning. IEEE Trans Affect Comput 12(2):544–550
74. Wen Z, Lin W, Wang T, Xu G (2021) Distract your attention: Multi-head cross attention network for facial expression recognition. arXiv preprint arXiv:2109.07270
75. Huang Q, Huang C, Wang X, Jiang F (2021) Facial expression recognition with grid-wise attention and visual transformer. Inf Sci 580:35–54
76. Greco A, Strisciuglio N, Vento M, Vigilante V (2022) Benchmarking deep networks for facial emotion recognition in the wild. Multimed Tools Appl. https://doi.org/10.1007/s11042-022-12790-7
77. Saurav S, Gidde P, Saini R, Singh S (2022) Dual integrated convolutional neural network for real-time facial expression recognition in the wild. Vis Comput 38(3):1083–1096
78. Saurav S, Saini AK, Saini R, Singh S (2022) Deep learning inspired intelligent embedded system for haptic rendering of facial emotions to the blind. Neural Comput Appl 34(6):4595–4623
79. Shi J, Zhu S, Liang Z (2021) Learning to amend facial expression representation via de-albino and affinity. arXiv preprint arXiv:2103.10189
80. Lundqvist D, Flykt A, Öhman A (1998) The karolinska directed emotional faces (kdef). CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet 91(630):2–2
81. da Silva FAM, Pedrini H (2015) Effects of cultural characteristics on building an emotion classifier through facial expression analysis. J Electron Imaging 24(2):023015
82. Xie S, Hu H (2018) Facial expression recognition using hierarchical features with deep comprehensive multipatches aggregation convolutional neural networks. IEEE Trans Multimed 21(1):211–220

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