


Article

# Facial Emotion Recognition Using Hybrid Features

Abdulrahman Alreshidi <sup>1</sup>  and Mohib Ullah <sup>2,\*</sup>

<sup>1</sup> College of Computer Science and Engineering, University of Ha'il, Ha'il 2440, Saudi Arabia; ab.alreshidi@uoh.edu.sa

<sup>2</sup> Department of Computer Science, Norwegian University of Science and Technology, 2815 Gjøvik, Norway

\* Correspondence: mohib.ullah@ntnu.no; Tel.: +47-405-757-21

Received: 10 January 2020; Accepted: 12 February 2020; Published: 18 February 2020



**Abstract:** Facial emotion recognition is a crucial task for human-computer interaction, autonomous vehicles, and a multitude of multimedia applications. In this paper, we propose a modular framework for human facial emotions' recognition. The framework consists of two machine learning algorithms (for detection and classification) that could be trained offline for real-time applications. Initially, we detect faces in the images by exploring the AdaBoost cascade classifiers. We then extract neighborhood difference features (NDF), which represent the features of a face based on localized appearance information. The NDF models different patterns based on the relationships between neighboring regions themselves instead of considering only intensity information. The study is focused on the seven most important facial expressions that are extensively used in day-to-day life. However, due to the modular design of the framework, it can be extended to classify  $N$  number of facial expressions. For facial expression classification, we train a random forest classifier with a latent emotional state that takes care of the mis-/false detection. Additionally, the proposed method is independent of gender and facial skin color for emotion recognition. Moreover, due to the intrinsic design of NDF, the proposed method is illumination and orientation invariant. We evaluate our method on different benchmark datasets and compare it with five reference methods. In terms of accuracy, the proposed method gives 13% and 24% better results than the reference methods on the static facial expressions in the wild (SFEW) and real-world affective faces (RAF) datasets, respectively.

**Keywords:** emotion recognition; cascade classifier; face detection; neighborhood difference features; random forest

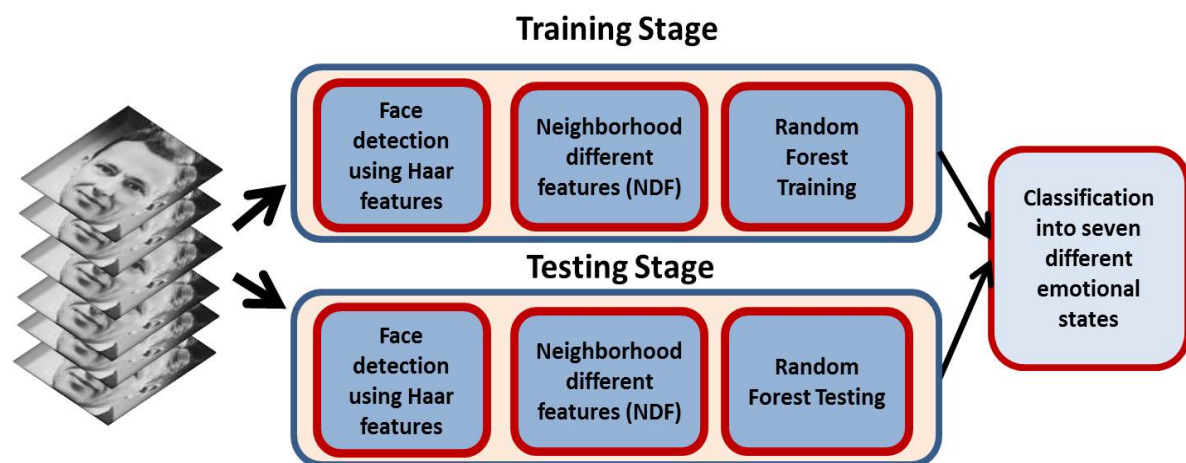
## 1. Introduction

Human emotion classification is described as a process to identify human emotion through facial expressions, verbal expressions, gestures and body movements, and multiple physiological signals' measurement. The significance of human feelings in the development of the latest technology gadgets is undeniable. In today's world, the analysis and recognition of emotion have an extensive range of importance in human-computer interaction, automated tutoring systems, image and video retrieval, smart environments, and driver warning systems [1]. In addition to that, emotion recognition plays a key role in determining various mental health conditions by psychiatrists and psychologists. In the past few decades, scientists and researchers have proposed different algorithms and techniques to recognize emotions from facial features and speech signals. It is still a challenging problem in the field of artificial intelligence, computer vision, psychology, and physiology due to the nature of its complexity. Scientists and researchers agree that facial expressions are the most influential part of recognizing human emotion. However, it is difficult to interpret humans' emotion by utilizing facial expression characteristics due to the sensitivity of the external noises, for example illumination conditions and dynamic head motion [2]. Moreover, the final results for emotion classification based on facial expressions still need to be improved.

To cope with these challenges, we propose a modular framework for emotions classification using only visual features. Our method does not rely on the assumption of a particular gender or skin color. As the neighborhood difference features encode the relative information of pixels with respect to their surroundings [3,4], our method is robust to illumination variations and face orientations. We mainly focus on seven facial emotions, i.e., anger, disgust, fear, happy, neutral, sad, and surprise. In a nutshell, the key contributions of the proposed framework are two-fold.

- The proposed framework is characterized by the compact representation of spatial information [3] that effectively encodes emotion information. The strengths of both face detection and emotion classification are integrated into a unified framework. The framework summarizes the local structure of the facial image considering the mutual relationship of neighboring regions.
- The framework is modular in nature based on three components including face detection, visual representation, and emotion classification. Therefore, any individual component can be replaced with any of the latest state-of-the-art algorithms. Furthermore, the individual components can be trained offline. Hence, the framework is suitable for handheld devices including smartphones.

The pipeline of our proposed method is depicted in Figure 1. In the rest of the paper, we divide the state-of-the-art methods into three categories in Section 2. We present our proposed method in Section 3 for the classification of facial emotions into the seven different classes. Experiments and results are presented in Section 4, and the conclusion is presented in Section 5.



**Figure 1.** Emotions' classification. The proposed method detects faces in the input frames/images and then extracts neighboring difference features (NDF) from the detected faces. These features are used to train a random forest classifier during the training stage. The same features are used to classify facial emotions into seven different states during the testing stage.

## 2. Related Work

We divided the state-of-the-art methods into three categories to elaborate the emotion classification techniques based on speech signals, physiological signals' measurements, and visual expressions.

### 2.1. Speech Signal Based Emotion Classification

Speech is a complex signal composed of numerous details. For example, it consists of information about the message to be communicated, speaker, language, region, and emotions. Speech processing can be considered as a major branch in digital signal processing. Additionally, speech processing has various applications in human-computer interfaces, telecommunication, assistive technology, and security. Likitha et al. [5] exploited the Mel frequency cepstral coefficient (MFCC) technique for emotion recognition through speech signals. Lotfidereshgi et al. [6] proposed a method based on

the combination of the classical source-filter model [7] and liquid state machine [8]. Deng et al. [9] proposed an emotion classification method associated with music. For this purpose, they investigated a three-dimensional resonance-arousal-valence model. However, the use of music stimuli alone may not be sufficient to study the dynamics of facial expressions. Tzirakis et al. [10] integrated a convolutional neural network with long short-term memory for speech emotion recognition. Considering multiple sources of speeches, the rate of recognizing an individual emotion will decline due to interference. To solve this problem, Sun et al. [11] proposed a speech emotion recognition method based on the decision tree support vector machine (SVM) with the Fisher feature selection model. Liu et al. [12] presented a speech emotion recognition method based on an improved brain emotional learning (BEL) model. Their model was inspired by the emotional processing mechanism of the limbic system in the brain. Speech based methods do not present good performance due to the lack of interaction among humans and machines. However, the speech signal can be used as a complementary source to improve the performance. For this purpose, cross-modality data can be used, including information extracted from physiological signals and image/video frames.

### *2.2. Physiological Signal Based Emotion Classification*

Physiological signals are obtained through different measurement methods. These methods include heartbeat rate (electrocardiogram or ECG/EKG signal), respiratory rate and content (capnogram), skin conductance (electrothermal activity or EDAsignal), muscle current (electromyography or EMG signal), and brain electrical activity (electroencephalography or EEG signal). These signals assist in determining the emotion of human beings. Ferdinando et al. [13] exploited multiple feature integration methods for emotion recognition based on ECG signals. Kanwal et al. [14] proposed a deep learning method for classifying different sleep stages using EEG signals. The framework has potential applications in the diagnosis of physiological and sleep-related disorders. Kanjo et al. [15] eliminated the need for manual feature extraction by exploiting multiple learning algorithms. They also consider the convolutional neural network and long short-term memory recurrent neural network based on information from phones and wearable devices. Nakisa et al. [16] solved the high-dimensionality problem of EEG signals by proposing a new framework to search for the optimal subset of EEG features automatically. They used evolutionary computation algorithms. For signal pre-processing and emotion classification, their method classified a wider set of emotions and integrated additional features. Ray et al. [17] presented an algorithm based on computational intelligence techniques for the analysis of EEG signals [18]. Vallverdù et al. [19] proposed a bio-inspired algorithm for emotion recognition. Jirayucharoensak et al. [20] proposed the implementation of a deep learning network to discover unknown feature correlations between input signals. Their method exploited a stacked autoencoder. However, these physiological signal based methods for emotion classification suffer from many problems [21]. These problems are the obtrusiveness of physiological sensors, the unreliability of physiological sensors, bodily position, air temperature, and humidity. Moreover, these signals have many-to-many relationship problems. In fact, multiple physiological signals can partially serve as indicators for multiple traditional biometric features. These signals also present varying time windows.

### *2.3. Visual Signal Based Emotion Classification*

Facial expression based emotion classification enhances the fluency, accuracy, and genuineness of an interaction. This type of classification approach is very useful in interpreting human-computer interaction [22]. Therefore, researchers are paying significant attention to facial expression based emotion classification methods. Sariyanidi et al. [4] showed that facial emotion recognition methods are driven by three components including face registration, representation, and finally, the recognition of different emotions. They also discussed different challenges, namely illumination and pose variations. For example, low-level texture features are helpful for the illumination invariant representation. For head pose variations, a part based model gives better results. In the case of video streams,

temporal information provides great assistance in the classification of different facial attributes. Jain et al. [23] proposed a model based on a single deep convolutional neural network comprised of convolution layers and deep residual blocks. Ullah et al. [24] trained a deep Siamese neural network with a contrastive loss to calculate the difference between image patches. The network has potential applications in tracking, image retrieval, and facial emotion recognition. Compared to deep learning based methods, Jeong et al. [25] proposed a simple machine learning based facial emotion recognition method. Acharya et al. [26] argued that regional distortion of facial images is useful for facial expressions. Based on this assumption, they trained a manifold network on top of a convolutional neural network to learn a variety of facial distortions. Rather than extracting only facial features, Ullah et al. [27] considered two set of low level features for inferring the human action. Wang et al. [28] used wavelet entropy and a single hidden layer feed-forward neural network for facial emotion classification. Yan [29] presented a collaborative discriminative multi-metric learning method for facial expression recognition considering videos. Multiple feature descriptors were computed for facial appearance and motion information from various aspects. Multiple distance metrics were then learned with these features to exploit complementary and discriminative information for recognition. Samadiani et al. [30] analyzed different modalities including EEG, infrared thermal, and facial sensors for emotion classification. They found that emotion classification was very susceptible to head orientation and illumination variation. Sun et al. [31] proposed a multi-channel deep neural network that learned and fused the spatial-temporal features. Lopes et al. [32] proposed several pre-processing steps before feeding the faces to a convolutional neural network for facial expression recognition. Franzoni et al. [33] took a step toward animal welfare and proposed a technique for recognizing the emotions of dogs. The proposed method has potential applications in supportive systems for pets. Another work fine-tuned the AlexNet [34] model through transfer learning on a dog dataset. Chen et al. [35] presented the histogram of oriented gradients from three orthogonal planes to extract dynamic textures from video sequences. To characterize facial appearance changes, a new effective geometric feature [36] was obtained from the warp transformation of facial landmarks. Researchers also focused on facial expression based emotion classification methods for handheld devices. In fact, smartphones and smart watches are equipped with a variety of sensors, which include accelerometer, gyroscope, fingerprint sensor, heart rate sensor, and microphone. Alshamsi et al. [37] proposed a framework that consisted of smartphone sensor technology supported by cloud computing for the real-time recognition of emotion in speech and facial expression. Hossain et al. [38] combined the potential of emotion-aware big data and cloud technology towards 5G. They combined facial and verbal features to present a bimodal system for big data emotion recognition. Grünerbl et al. [39] introduced a system based on smartphone sensors for the recognition of depressive and manic states. They detected state changes of patients suffering from bipolar disorder. Sneha et al. [40] analyzed the textual content of the message and user typing behavior to classify future instances. Hossain et al. [41] modeled hybrid features based on bandelet transform [42] and local binary patterns [43] for emotion classification. They performed classification based on a Gaussian mixture model. Sokolov et al. [44] proposed a cross-platform application for emotion recognition. Their application was based on a convolutional neural network and was capable of recognizing emotions on the arousal-valence scale. Perikos et al. [45] considered shape deformation for eyebrows, mouth, eyes, and lips to extract the features. These features were then used for the facial emotion recognition. Franzoni et al. [46] analyzed the semantic proximity of sentences to analyze emotional content.

The literature is very limited due to the associated challenges of developing a reliable technique with low computational requirements. The aforementioned methods required huge computational power since most of them were based on deep models. These methods were modeled for very narrow and specific emotions, and they were not easily expandable to consider other emotional states. Compared to that, the design of our proposed framework is modular and can be extended easily to classify complex facial emotions. Moreover, our proposed method presents a compact representation

of spatial information [3] that effectively encodes emotion information. We integrate the strengths of both face detection and emotion classification into a unified model.

### 3. Proposed Method

Feature modeling and classification accuracy are strongly related to each other. We modeled robust features that filter out redundant and irrelevant information from detected faces. Additionally, the features were illumination and orientation invariant. For classification, we used random forest, which is a powerful and accurate classifier. Random forest presents good performance on many problems including non-linear problems. Due to the classification strengths of the trees in the random forest, our method avoided both overfitting and underfitting. Random forest renders good performance by training it even with small samples. Considering our proposed features, this made the classifier ideal for different personality traits and high segmented facial expression. The random forest presents generalization capability. Therefore, our proposed method could handle unseen data. The generalization capability of our method was determined by the complexity and training of the random forest. Our proposed framework was easily expandable to classify more complex facial emotions. It was only a matter of the training stage regardless of the number of emotional states. Our proposed method could be divided into three stages/modules that are listed as:

- Face detection,
- Extraction of NDF features,
- Emotion state classification.

#### 3.1. Face Detection

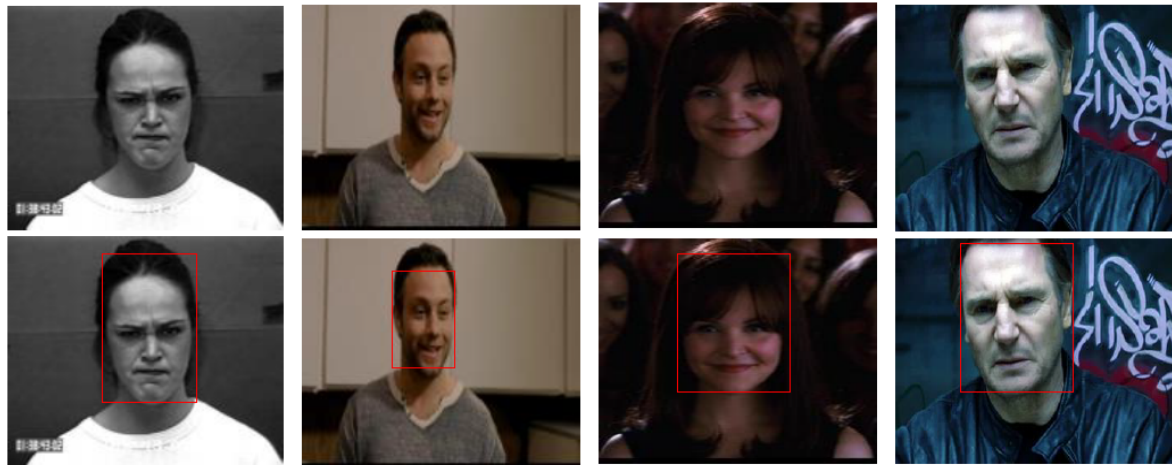
The Haar-like feature based cascade method is a famous face detection model [47,48] due to its simplicity and robustness. Inspired by the model, we trained a cascade function considering ground truth faces with their labels. The model entailed positive labels for faces and negative labels for non-faces to train the classifier. Subsequently, we extracted Haar features from the detected faces (Figure 2) that resembled convolutional kernels. Each feature was a single value calculated by subtracting a rectangular region from another region in the same frame. Due to different rectangles, we exploited different sizes and locations of each kernel to obtain many features. For this purpose, we exploited the concept of the integral image [49].

$$\begin{aligned}\rho(x, y) &= \sum_{x' \leq x, y' \leq y} \psi(x', y') \\ \sigma(x, y) &= \sigma(x, y - 1) + \psi(x, y) \\ \rho(x, y) &= \rho(x - 1, y) + \sigma(x, y)\end{aligned}\tag{1}$$

where  $\rho$  is the integral image and  $\psi(x', y')$  is the original image.  $\sigma(x, y)$  is the cumulative row sum. The integral image could be obtained in one pass over the original image. Additionally, we explored the AdaBoost model [50] to filter out irrelevant features. To remove irrelevant features, we considered each and every feature on all the training images. For each feature, we investigated the optimal threshold that would classify faces and non-faces. We chose the features with the smallest error rate since these features classified the faces and non-faces in an optimized way. In the beginning, each image was given an equal weight. After each classification, we increased the weights of misclassified images and repeated the same procedure. We then calculated new error rates and new weights.

We observed that in each image, the major part consisted of the non-face region. Therefore, if a window was not comprised of a face, we filtered it out in the subsequent stage of classification through the latent emotional state. To reduce the number of misdetections and false positives, we exploited the concept of the cascade of classifiers [51]. We combined the features into different stages of classifiers and explored them one-by-one. We removed the window if it did not qualify in the first stage. Therefore, we did not explore the remaining features. If the window qualified the first stage, we applied the

second stage of features and continued the procedure. A window that qualified all stages was a face region.



**Figure 2.** Face detection results of the trained model. Only the detected bounding box region of the whole frame is used to classify the facial emotion.

### 3.2. Neighborhood Difference Features

After localizing the face in a given image, we then extracted neighborhood difference features (NDF). The extracted NDF was represented by localized appearance information. Our NDF formulated different patterns based on the relationships between neighboring regions. For appearance information, we explored the neighborhood in numerous directions and scaled to compute regional patterns. We determined the correspondence between neighboring regions by using the extrema on appearance values. We wanted to summarize efficiently the local structures of the face by exploiting each pixel as a center pixel in a region. Considering a center pixel  $S_c$  and neighboring pixels  $S_n$  ( $n = 1, 2, \dots, 8$ ) in a detected face, we computed the pattern number (PN) as,

$$(PN)_{M,N} = \sum_{n=0}^{M-1} 2^n x \tau_1(S_n - S_c) \tag{2}$$

$$\tau_1 = \begin{cases} 1, & \text{if } (S_n - S_c) > 0 \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

where  $M$  and  $N$  are the radii of neighbors and the number of neighbors for the pattern number. After calculating the  $PN$  of a face, a histogram is computed as,

$$\pi_1(l) = \sum_{x=1}^M \sum_{y=1}^N \tau_2(PN_{x,y}, l) : l \in [0, 2^M - 1] \tag{4}$$

$$\tau_2(a, b) = \begin{cases} 1, & \text{if } \mathbf{a} = \mathbf{b} \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

The relationship between regions in terms of these pixels was exploited, and a pattern number was assigned. We constructed a histogram to represent the face in the form of NDF. For neighboring pixels  $S_n$  and a center pixel  $S_c$ , NDF could be formulated as,

$$\begin{aligned}
G_1^n &= S_8 - S_n, G_2^n = S_{n+1} - S_n, \text{ for } n = 1 \\
G_1^n &= S_{n-1} - S_n, G_2^n = S_{n+1} - S_n, \forall n = 1, 2, \dots, 8 \\
G_1^n &= S_{n-1} - S_n, G_2^n = S_1 - S_n, \text{ for } n = 8
\end{aligned} \tag{6}$$

We found the difference of each neighbor with two other neighbors in  $G_1^n$  and  $G_2^n$ . Considering these two differences, we assigned a pattern number to each neighbor,

$$\tau_3(G_1^n - G_2^n) = \begin{cases} 1, & \text{if } G_1^n \geq 0, \text{ and } G_2^n \geq 0 \\ 1, & \text{if } G_1^n < 0, \text{ and } G_2^n < 0 \\ 0, & \text{if } G_1^n \geq 0, \text{ and } G_2^n < 0 \\ 0, & \text{if } G_1^n < 0, \text{ and } G_2^n \geq 0 \end{cases} \tag{7}$$

For the central pixel  $S_c$ , NDF could be found using the above numbers, and the histogram for NDF map could be calculated in the equations,

$$\begin{aligned}
NDF(S_c) &= \sum_{n=1}^8 2^{n-1} x \tau_3(G_1^n - G_2^n) \\
\pi_2(NDF) &= \sum_{x=1}^M \sum_{y=1}^N \tau_2(NDF_{x,y}, l) : l \in [0, 2^8 - 1]
\end{aligned} \tag{8}$$

The NDF represented novel features that were obtained by extracting the relationship among neighboring regions by considering them mutually. The NDF computed the relationship of neighboring regions with the central region. In the proposed framework, face detection and NDF worked sequentially as they competed with each other based on the characteristics they represented individually.

### 3.3. Emotion Classification

To classify NDF features into the corresponding emotional class, we explored the random forest classifier (RFC) [52,53]. The RFC consisted of random trees, which were a combination of predictors. The RFC took the input features and classified them with every tree in the classifier. It then provided the class label that obtained the majority of votes. The classifier was trained with the same parameters considering the training sets. These sets were produced from the original training set using the bootstrap process. For each training set, the classifier chose the same number of features as in the original set. The features were selected with replacement. This meant that some features would be taken more than once and some would be negligible. At each node of each trained tree, the classifier used a subset of variables to determine the best split. With each node, a new subset was produced. The random classifier did not require cross-validation or bootstrapping or a separate test set to get an approximation of the training error. The error was computed internally during the training. When the training set for the current tree was drawn by sampling with replacement, some features were left out, which were called out-of-bag (OOB) data. The classification error was computed by exploring these OOB data. For this purpose, the classifier obtained a prediction for each feature, which was OOB relative to the  $i^{\text{th}}$  tree.

## 4. Experiments and Results

For the experimental evaluation, we considered the static facial expressions in the wild (SFEW) 2.0 dataset [54,55] and the real-world affective faces (RAF) dataset [56]. To train the RFC considering seven emotions, namely anger, disgust, fear, happy, neutral, sad, and surprise, a latent emotion was

introduced. We assumed that when the face was detected correctly in a given image, it would be classified into either of the seven classes. If the face was not detected correctly, it would be assigned to the latent emotion. Therefore, the misclassified faces were handled and their impact on the final error was minimized. We performed a comparison with five state-of-the-art methods and reported the results in terms of confusion matrices, precision-recall, and the total accuracies.

#### 4.1. Performance Metric

We evaluated the performance of our framework using confusion matrices, precision-recall, and the total accuracy. We calculated a multi-class confusion matrix as shown in Table 1 and Table 2. We calculated recall and precision as provided in the equations:

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

where TP, FP, and FN represent true positive, false positive, and false negative, respectively. For the multi-class confusion matrix, the average accuracy is calculated as:

$$AverageAccuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

where TP, TN, FP, and FN are the overall true positive, true negative, false positive, and false negative of all the classes in the confusion matrix. In other words, the overall accuracy was the sum of off-diagonal elements divided by all the elements in the multi-class confusion matrix. The proposed framework was compared to five state-of-the-art methods.

**Table 1.** Confusion matrix for the static facial expressions in the wild (SFEW) dataset. The results for seven emotions are presented.

		Predicted Facial Emotion							Recall
		Anger	Disgust	Fear	Happy	Neutral	Sad	Surprise	
Actual Facial Emotion	Anger	0.68	0.02	0.10	0.04	0.05	0.06	0.05	68%
	Disgust	0.20	0.55	0.04	0.03	0.01	0.01	0.16	55%
	Fear	0.10	0.33	0.49	0.02	0.04	0.01	0.01	49%
	Happy	0.02	0.30	0.01	0.51	0.01	0.11	0.04	51%
	Neutral	0.05	0.19	0.02	0.10	0.52	0.03	0.09	52%
	Sad	0.11	0.05	0.07	0.01	0.12	0.62	0.02	62%
	Surprise	0.13	0.06	0.01	0.10	0.02	0.01	0.67	67%
Precision		52.71%	36.66%	66.21%	62.96%	67.53	72.94%	64.42%	



**Table 2.** Confusion matrix for the real-world affective faces (RAF) dataset.

		Predicted Facial Emotion							Recall
		Anger	Disgust	Fear	Happy	Neutral	Sad	Surprise	
Actual Facial Emotion	Anger	0.54	0.17	0.06	0.01	0.03	0.15	0.04	54%
	Disgust	0.11	0.57	0.15	0.02	0.01	0.01	0.13	57%
	Fear	0.12	0.20	0.61	0.03	0.03	0.01	0.00	61%
	Happy	0.30	0.04	0.01	0.49	0.05	0.04	0.07	49%
	Neutral	0.10	0.06	0.13	0.02	0.65	0.01	0.03	65%
	Sad	0.01	0.01	0.02	0.20	0.00	0.69	0.07	69%
	Surprise	0.05	0.03	0.01	0.15	0.08	0.10	0.58	58%
Precision		43.90%	52.77%	61.61%	53.26%	76.47%	68.31%	63.04%	

It is worth noting that the anger facial expression is a tense emotional outcome. It happens when a person considers that his/her personal limits are violated. Persons in this kind of emotion generally perform gestures including intense staring with the eyes wide open, output uncomfortable sounds, bare the teeth, and attempt to physically seem larger. Staring with the eyes wide open is a significant hint for computers to recognize anger. There are also other face-related elements including V-shaped eyebrows, wrinkled nose, narrowed eyes, and forwarded jaws. All these important elements help to recognize anger emotion. In facial expression, happy indicates an emotional state of joy. In this emotional state, the reader can find that the forehead muscle relaxes and the eyebrows are pulled up slowly. Apart from that, both the wrinkled outer corners of eyes and pulled up lip corners represent unique representations. In fact, the neutral facial emotion relaxes the muscles of the face. Other facial emotions need to use extensive muscles of the face.

From Tables 1 and 2, two general findings are observed. Firstly, anger and disgust interfered with each other easily. Secondly, fear and surprise interfered with each other easily. These observations could be the early phase of dynamic facial expression between anger and disgust. They resembled each other due to the similar and aligned movement of the nose wrinkle and lip funneler. Considering both fear and surprise, the upper lid raise and jaw drop may contribute to the same conclusion.

We also compared our proposed method with five reference methods over two datasets. These reference methods included the implicit fusion model [57], bi-orthogonal model [58], higher order model [59], bio-inspired model [60], and statistical similarity model [3]. The comparison results are listed in Table 3 in terms of the total accuracies. Our proposed method achieved promising results and performed better than the five reference methods. The average accuracy of our proposed framework gave 13% and 24% better results on the two datasets compared to the reference methods. Our method still had some limitations. For example, we did not exploit geometric features, which could complement the performance. Our method is applicable to treating and diagnosing patients with emotional issues.

**Table 3.** Total accuracies are presented for the reference methods and our proposed method considering both datasets, namely SFEW and RAF.

Methods	SFEW Total Accuracy	RAF Total Accuracy
Han et al. [57]	56.4	55.7
Zhang et al. [58]	54.2	56.6
Ali et al. [59]	49.8	52.7
Vivek et al. [60]	53.5	54.9
Verma et al. [3]	47.6	48.1
Prop. method	57.7	59.0

## 5. Conclusions

We presented a modular framework for facial emotion classification into seven different states. For this purpose, we detected faces and extracted neighborhood difference features (NDF) based on the relationships between neighboring regions. To classify facial emotions into seven different classes, we trained a random forest classifier that recognized these emotions during the testing stage. We evaluated our method on two benchmark datasets and compared it with five reference methods where our method outperformed on both datasets. In our future work, we will extend our method to videos. For videos, we will investigate and combine both spatial and temporal features into a unified model. Moreover, geometric features and facial deformation will be explored and integrated into the proposed framework.

**Author Contributions:** A.A. is the first author of the paper. His contributions consist of methodology, data collection, implementation, and testing. M.U. did paper writing and correspondence with the reviewers. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

NDF	Neighborhood difference features
LSM	Liquid state machine
CNN	Convolutional neural network
BEL	Brain emotional learning
SNN	Spiking neural network
LDA	Linear discriminant analysis
MCML	Maximally collapsing metric learning
BWT	Bionic wavelet transform
MFCC	Mel frequency cepstral coefficient

## References

1. Shojailangari, S.; Yau, W.Y.; Nandakumar, K.; Li, J.; Teoh, E.K. Robust representation and recognition of facial emotions using extreme sparse learning. *IEEE Trans. Image Process.* **2015**, *24*, 2140–2152. [[CrossRef](#)]
2. Ko, K.E.; Sim, K.B. Development of a Facial Emotion Recognition Method based on combining AAM with DBN. In Proceedings of the 2010 International Conference on Cyberworlds (CW), Singapore, 20–22 October 2010; pp. 87–91.
3. Verma, M.; Raman, B. Local neighborhood difference pattern: A new feature descriptor for natural and texture image retrieval. *Multimed. Tools Appl.* **2018**, *77*, 11843–11866. [[CrossRef](#)]
4. Sariyanidi, E.; Gunes, H.; Cavallaro, A. Automatic analysis of facial affect: A survey of registration, representation, and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *37*, 1113–1133. [[CrossRef](#)] [[PubMed](#)]
5. Likitha, M.; Gupta, S.R.R.; Hasitha, K.; Raju, A.U. Speech based human emotion recognition using MFCC. In Proceedings of the 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, India, 20–24 March 2017; pp. 2257–2260.
6. Lotfidereshgi, R.; Gournay, P. Biologically inspired speech emotion recognition. In Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 5–9 March 2017; pp. 5135–5139.
7. Durrieu, J.L.; Richard, G.; David, B.; Févotte, C. Source/filter model for unsupervised main Melody extraction from polyphonic audio signals. *IEEE Trans. Audio Speech Lang. Process.* **2010**, *18*, 564–575. [[CrossRef](#)]
8. Verstraeten, D.; Schrauwen, B.; Stroobandt, D.; Van Campenhout, J. Isolated word recognition with the liquid state machine: A case study. *Inf. Process. Lett.* **2005**, *95*, 521–528. [[CrossRef](#)]

9. Deng, J.J.; Leung, C.H.; Milani, A.; Chen, L. Emotional states associated with music: Classification, prediction of changes, and consideration in recommendation. *ACM Trans. Interact. Intell. Syst. TiiS* **2015**, *5*, 4. [[CrossRef](#)]
10. Tzirakis, P.; Zhang, J.; Schuller, B.W. End-to-end speech emotion recognition using deep neural networks. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 5089–5093.
11. Sun, L.; Fu, S.; Wang, F. Decision tree SVM model with Fisher feature selection for speech emotion recognition. *EURASIP J. Audio, Speech Music Process.* **2019**, *2019*, 2. [[CrossRef](#)]
12. Liu, Z.T.; Xie, Q.; Wu, M.; Cao, W.H.; Mei, Y.; Mao, J.W. Speech emotion recognition based on an improved brain emotion learning model. *Neurocomputing* **2018**, *309*, 145–156. [[CrossRef](#)]
13. Ferdinando, H.; Seppänen, T.; Alasaarela, E. Enhancing Emotion Recognition from ECG Signals using Supervised Dimensionality Reduction. In Proceedings of the ICPRAM, Porto, Portugal, 24–26 February 2017; pp. 112–118.
14. Kanwal, S.; Uzair, M.; Ullah, H.; Khan, S.D.; Ullah, M.; Cheikh, F.A. An Image Based Prediction Model for Sleep Stage Identification. In Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 22–25 September 2019; pp. 1366–1370.
15. Kanjo, E.; Younis, E.M.; Ang, C.S. Deep learning analysis of mobile physiological, environmental and location sensor data for emotion detection. *Inf. Fusion* **2019**, *49*, 46–56. [[CrossRef](#)]
16. Nakisa, B.; Rastgoo, M.N.; Tjondronegoro, D.; Chandran, V. Evolutionary computation algorithms for feature selection of EEG-based emotion recognition using mobile sensors. *Expert Syst. Appl.* **2018**, *93*, 143–155. [[CrossRef](#)]
17. Ray, P.; Mishra, D.P. Analysis of EEG Signals for Emotion Recognition Using Different Computational Intelligence Techniques. In *Applications of Artificial Intelligence Techniques in Engineering*; Springer: Berlin, Germany, 2019; pp. 527–536.
18. Ullah, H.; Uzair, M.; Mahmood, A.; Ullah, M.; Khan, S.D.; Cheikh, F.A. Internal emotion classification using eeg signal with sparse discriminative ensemble. *IEEE Access* **2019**, *7*, 40144–40153. [[CrossRef](#)]
19. Franzoni, V.; Vallverdù, J.; Milani, A. Errors, Biases and Overconfidence in Artificial Emotional Modeling. In Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence-Companion Volume, Thessaloniki, Greece, 14–17 October 2019; pp. 86–90.
20. Jirayucharoensak, S.; Pan-Ngum, S.; Israsena, P. EEG-based emotion recognition using deep learning network with principal component based covariate shift adaptation. *Sci. World J.* **2014**, *2014*, 627892. [[CrossRef](#)] [[PubMed](#)]
21. van den Broek, E.L.; Spitters, M. Physiological signals: The next generation authentication and identification methods!? In Proceedings of the 2013 European Intelligence and Security Informatics Conference (EISIC), Uppsala, Sweden, 12–14 August 2013; pp. 159–162.
22. Rota, P.; Ullah, H.; Conci, N.; Sebe, N.; De Natale, F.G. Particles cross-influence for entity grouping. In Proceedings of the 21st European Signal Processing Conference (EUSIPCO 2013), Marrakech, Morocco, 9–13 September 2013; pp. 1–5.
23. Jain, D.K.; Shamsolmoali, P.; Sehdev, P. Extended Deep Neural Network for Facial Emotion Recognition. *Pattern Recognit. Lett.* **2019**, *120*, 69–74. [[CrossRef](#)]
24. Ullah, M.; Ullah, H.; Cheikh, F.A. Single shot appearance model (ssam) for multi-target tracking. *Electron. Imaging* **2019**, *2019*, 466-1–466-6. [[CrossRef](#)]
25. Jeong, M.; Ko, B.C. Driver's Facial Expression Recognition in Real-Time for Safe Driving. *Sensors* **2018**, *18*, 4270. [[CrossRef](#)] [[PubMed](#)]
26. Acharya, D.; Huang, Z.; Pani Paudel, D.; Van Gool, L. Covariance pooling for facial expression recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 367–374.
27. Ullah, M.; Ullah, H.; Alseadonn, I.M. Human action recognition in videos using stable features. In Proceedings of the ICPRAM, Porto, Portugal, 24–26 February 2017.
28. Wang, S.H.; Phillips, P.; Dong, Z.C.; Zhang, Y.D. Intelligent facial emotion recognition based on stationary wavelet entropy and Jaya algorithm. *Neurocomputing* **2018**, *272*, 668–676. [[CrossRef](#)]
29. Yan, H. Collaborative discriminative multi-metric learning for facial expression recognition in video. *Pattern Recognit.* **2018**, *75*, 33–40. [[CrossRef](#)]

30. Samadiani, N.; Huang, G.; Cai, B.; Luo, W.; Chi, C.H.; Xiang, Y.; He, J. A review on automatic facial expression recognition systems assisted by multimodal sensor data. *Sensors* **2019**, *19*, 1863. [[CrossRef](#)]
31. Sun, N.; Li, Q.; Huan, R.; Liu, J.; Han, G. Deep spatial-temporal feature fusion for facial expression recognition in static images. *Pattern Recognit. Lett.* **2019**, *119*, 49–61. [[CrossRef](#)]
32. Lopes, A.T.; de Aguiar, E.; De Souza, A.F.; Oliveira-Santos, T. Facial expression recognition with convolutional neural networks: coping with few data and the training sample order. *Pattern Recognit.* **2017**, *61*, 610–628. [[CrossRef](#)]
33. Franzoni, V.; Milani, A.; Biondi, G.; Micheli, F. A Preliminary Work on Dog Emotion Recognition. In Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence-Companion Volume, Thessaloniki, Greece, 14–17 October 2019; pp. 91–96.
34. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*; Communications of the ACM: New York, NY, USA, 2012; pp. 1097–1105.
35. Chen, J.; Chen, Z.; Chi, Z.; Fu, H. Facial expression recognition in video with multiple feature fusion. *IEEE Trans. Affect. Comput.* **2018**, *9*, 38–50. [[CrossRef](#)]
36. Ullah, H.; Altamimi, A.B.; Uzair, M.; Ullah, M. Anomalous entities detection and localization in pedestrian flows. *Neurocomputing* **2018**, *290*, 74–86. [[CrossRef](#)]
37. Alshamsi, H.; Kepuska, V.; Alshamsi, H.; Meng, H. Automated Facial Expression and Speech Emotion Recognition App Development on Smart Phones using Cloud Computing. In Proceedings of the 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, Canada, 1–3 November 2018; pp. 730–738.
38. Hossain, M.S.; Muhammad, G.; Alhamid, M.F.; Song, B.; Al-Mutib, K. Audio-visual emotion recognition using big data towards 5G. *Mob. Netw. Appl.* **2016**, *21*, 753–763. [[CrossRef](#)]
39. Grünerbl, A.; Muaremi, A.; Osmani, V.; Bahle, G.; Oehler, S.; Tröster, G.; Mayora, O.; Haring, C.; Lukowicz, P. Smartphone-based recognition of states and state changes in bipolar disorder patients. *IEEE J. Biomed. Health Inform.* **2015**, *19*, 140–148. [[CrossRef](#)]
40. Sneha, H.; Rafi, M.; Kumar, M.M.; Thomas, L.; Annappa, B. Smartphone based emotion recognition and classification. In Proceedings of the 2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, Tamil Nadu, India, 22–24 February 2017; pp. 1–7.
41. Hossain, M.S.; Muhammad, G. An emotion recognition system for mobile applications. *IEEE Access* **2017**, *5*, 2281–2287. [[CrossRef](#)]
42. Mosleh, A.; Bouguila, N.; Hamza, A.B. Video completion using bandelet transform. *IEEE Trans. Multimed.* **2012**, *14*, 1591–1601. [[CrossRef](#)]
43. Heikkilä, M.; Pietikäinen, M.; Schmid, C. Description of interest regions with local binary patterns. *Pattern Recognit.* **2009**, *42*, 425–436. [[CrossRef](#)]
44. Sokolov, D.; Patkin, M. Real-time emotion recognition on mobile devices. In Proceedings of the 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), Xi'an, China, 15–19 May 2018; p. 787.
45. Perikos, I.; Paraskevas, M.; Hatzilygeroudis, I. Facial expression recognition using adaptive neuro-fuzzy inference systems. In Proceedings of the 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS), Singapore, 6–8 June 2018; pp. 1–6.
46. Franzoni, V.; Biondi, G.; Milani, A. A web-based system for emotion vector extraction. In Proceedings of the International Conference on Computational Science and Its Applications, Trieste, Italy, 3–6 July 2017; pp. 653–668.
47. Aguilar, W.G.; Luna, M.A.; Moya, J.F.; Abad, V.; Parra, H.; Ruiz, H. Pedestrian detection for UAVs using cascade classifiers with meanshift. In Proceedings of the 2017 IEEE 11th International Conference on Semantic Computing (ICSC), San Diego, CA, USA, 30 January–1 February 2017; pp. 509–514.
48. Viola, P.; Jones, M. Rapid object detection using a boosted cascade of simple features. In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Kauai, HI, USA, 8–14 December 2001; Volume 1.
49. Bradley, D.; Roth, G. Adaptive thresholding using the integral image. *J. Graph. Tools* **2007**, *12*, 13–21. [[CrossRef](#)]
50. Hastie, T.; Rosset, S.; Zhu, J.; Zou, H. Multi-class adaboost. *Stat. Its Interface* **2009**, *2*, 349–360. [[CrossRef](#)]

51. Bruzzone, L.; Cossu, R. A multiple-cascade-classifier system for a robust and partially unsupervised updating of land-cover maps. *IEEE Trans. Geosci. Remote. Sens.* **2002**, *40*, 1984–1996. [[CrossRef](#)]
52. Cutler, A.; Cutler, D.R.; Stevens, J.R. Random forests. In *Ensemble Machine Learning*; Springer: Berlin, Germany, 2012; pp. 157–175.
53. Au, T.C. Random forests, decision trees, and categorical predictors: the Absent levels problem. *J. Mach. Learn. Res.* **2018**, *19*, 1737–1766.
54. Dhall, A.; Goecke, R.; Joshi, J.; Sikka, K.; Gedeon, T. Emotion recognition in the wild challenge 2014: Baseline, data and protocol. In Proceedings of the 16th International Conference on Multimodal Interaction, Istanbul, Turkey, 12–16 November 2014; pp. 461–466.
55. Dhall, A.; Goecke, R.; Lucey, S.; Gedeon, T. Collecting large, richly annotated facial-expression databases from movies. *IEEE Multimed.* **2012**, *19*, 34–41. [[CrossRef](#)]
56. Li, S.; Deng, W.; Du, J. Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2852–2861.
57. Han, J.; Zhang, Z.; Ren, Z.; Schuller, B. Implicit Fusion by Joint Audiovisual Training for Emotion Recognition in Mono Modality. In Proceedings of the ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 5861–5865.
58. Zhang, Y.D.; Yang, Z.J.; Lu, H.M.; Zhou, X.X.; Phillips, P.; Liu, Q.M.; Wang, S.H. Facial emotion recognition based on biorthogonal wavelet entropy, fuzzy support vector machine, and stratified cross validation. *IEEE Access* **2016**, *4*, 8375–8385. [[CrossRef](#)]
59. Ali, H.; Hariharan, M.; Yaacob, S.; Adom, A.H. Facial emotion recognition based on higher-order spectra using support vector machines. *J. Med. Imaging Health Inform.* **2015**, *5*, 1272–1277. [[CrossRef](#)]
60. Vivek, T.; Reddy, G.R.M. A hybrid bioinspired algorithm for facial emotion recognition using CSO-GA-PSO-SVM. In Proceedings of the 2015 Fifth International Conference on Communication Systems and Network Technologies, Gwalior, India, 4–6 April 2015; pp. 472–477.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).