International Journal of Informatics and Applied Mathematics
e-ISSN:2667-6990 Vol. 3, No. 1, 39-53

Automated Facial Expression Recognition Using Deep Learning Techniques: An Overview

Meriem Sari[1], Abdelouahab Moussaoui[1], and Abdenour Hadid[2]

[1] Department of Computer Science, University of Ferhat Abbas Setif1, Setif Algeria
{meriem.sari , abdelouahab.moussaoui}@univ-setif.dz
[2] Center for Machine Learning and Signal Analysis, University of Oulu, Oulu Finland
hadid.abdenour@oulu.fi

Abstract. Facial expression recognition (FER) plays a key role in conveying human emotions and feelings. Automated FER systems enable different machines to recognize emotions without the help of humans; this is considered as a very challenging problem in machine learning. Over the years there has been a considerable progress in this field. In this paper we present a state of the art overview on the different concepts of a FER system and the different used methods; plus we studied the efficiency of using deep learning architectures specifically convolutional neural networks architectures (CNN) as a new solution for FER problems by investigating the most recent and cited works.

Keywords: Facial expression recognition · Emotion Recognition · Machine learning · Deep learning · Convolutional Neural Network

1 Introduction

Facial expression is one of the most important aspects of biometry; it has been regarded as a fresh and active research field in the last decade due to its importance in translating the emotional state of peaple. Even though this latter analysis can be done through other features such as: voice [11], body gestures, social and contextual parameters of the situation [25] among others facial expression remains to be the most expressive way through which human beings can display their emotions because it has a high level of directness, friendliness, convenience and robustness.

Nowadays, facial expression recognition (FER) has known a large number of applications thanks to the huge amount of attention that it got. It is mostly used in human machine interaction (HCI) applications such as; interactive gaming, digital entertainment, virtual reality and robotics. It is also used in emotion and behavioural analysis in the medical domain (Autism [17], mental disorder [48], pain assessment [36]) also used in surveillance and law enforcement applications.

Facial expression was first introduced as a research field by Darwin in his book The Expression of the Emotions in Man and Animals [7], and then it was studied by many other scientists. In the recent few decades the work of Paul Ekman has been considered as the cornerstone of almost all the research done in this field [10]; Ekman introduced the six basic emotions [9] which are; happiness, sadness, anger, disgust, fear, surprise plus the neutral emotion that is considered in most of the works, these emotions became universal among human beings.

Automated FER System (AFERS or FERS) is the very complicated process for machines to automatically recognize emotions without any help from human beings. This system receives as input images that contain faces, it performs some processing that will be cited in the next section and then gives as output the recognized emotion. Even though it seems to be a very simple task for humans, it represents a very challenging one in the world of machine learning. In this paper we try firstly to give a brief summary on the FER concepts enriched by what scientists had accomplished over the years then we present a comparative study of the most recent works that have been done in the field using deep learning.

The rest of the paper is organized as follow: next section represents a background review about FER concepts. Section III represents a comprehensive study for FER problems using deep learning. Observations and discussion are presented in section IV then conclusion and future directions in section V.

2 Background Review

FERS can be divided into two main categories in terms of inputs; i.e. it can take images or dynamic sequences; in either cases; the FER system workflow is shown in Figure 1 , we observe that the analysis is done via three important steps:

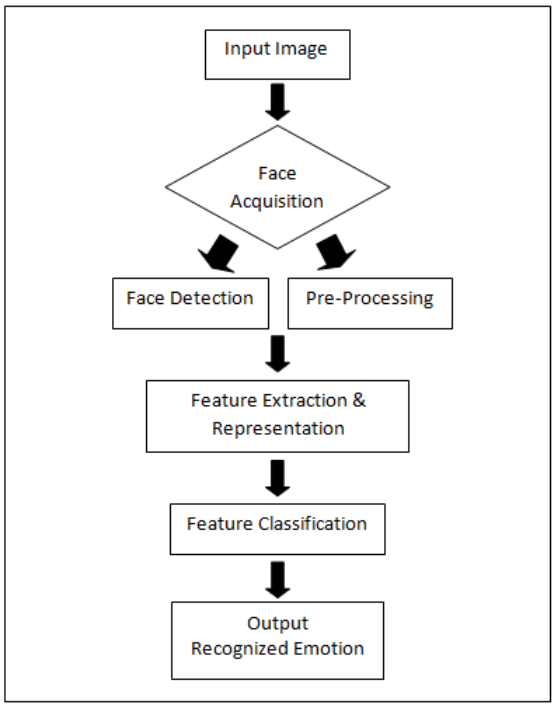

Fig. 1. Facial Expression Recognition System Workflow

2.1 Face Acquisition

Over the years there have been many methods developed to detect faces in an arbitrary scene [19, 31, 42, 43, 46, 49], some can detect only frontal view faces while others can work even on multi-view faces such as side views, the most common and used technique for face detection is the algorithm of Viola and Jones [55]. Over the time, this phase was assumed to be a pre-processing phase along with the modifications that we apply on different input images before we feed them to the FES. A pre-processing phase could also include image resizing, denoising, rotation correction, etc; it is based on suppressing unwanted distortions and enhancing image features in order to improve the quality of the input data that the system is going to work on.

2.2 Feature Extraction and Representation

After locating the face, the next phase is to extract information from the input data known as features. In this process variety of features can be retrieved and based on this we can categorize the used methods for feature extraction. Over the literature, the facial characteristics are divided into two main categories.

Geometric features and they represent the shape and location of facial components (mouth, eyes, eyebrows, nose) during the execution of a facial expression.

The best way to describe nearly the entire facial component movements is the Facial Action Coding System (FACS) introduced by Ekman [13], it contains 44 Action Units (AU) and another 20 was added in 2002 [10]; each AU describes a set of facial muscles that works together to perform the movements related to a specific facial expression. Methods that are based on geometry use a feature vector that is composed of facial feature points after their extraction, some examples of application of this method can be found in [14, 44, 60, 16, 27, 51].

On the other hand Appearance Features represent the skin or texture changes of the face without taking in consideration the muscles motion. Methods based on appearance features are considered time consuming in term of processing yet studies have shown that they produce better results because they take into account pixel intensity, texture edges, color arrangements, wrinkles and furrows of the whole face or just a small part of it. The most successful methods using appearance features are the local binary pattern (LBP) and its extensions [1, 26]. For further examples we can refer to [51, 57, 16, 24, 20].

2.3 Feature Classification and Emotion Recognition

The previous phase results a large set of features, the aim of this phase is to choose only the most discriminative ones in order to decrease the processing time and give more credibility to the results. Over the years, researchers have developed and innovated a considerable number of methods; some of the very known are: Support Vector Machines (SVM) [62], Bayes classifier [6], Fuzzy Techniques [3], Feature Selection [8], Artificial Neural Networks (ANN) [23] and others [64]. The general workflow of a FER system is presented in

Despite the increasing improvements in the recognition performances, there are still some limitations in terms of accuracy (subjectivity, occlusion, pose, low resolution, scale, illumination variation, spontaneous vs. posed expressions) and efficiency (time, computational and space complexity) [51, 54, 53]. To overcome those limitations researchers started digging to find new methods or focused on enhancing existing ones that lead to using Deep Neural Networks (DNN) which are based on regular ANN.

2.4 Deep Neural Networks

Artificial Neural Networks are based on the idea of mathematically representing information processing in human brains, most of them rely on features selected or created by humans whereas Deep Neural Networks (DNN) have the exact same structure as classical ANN, the only difference is that they use multiple and deeper hidden layers instead of just one and that gives more chance to increasingly learn bigger and more complex and abstract features. The crucial point of DNN is that they create a hierarchy of representations, each more and more abstract than the preceding to ultimately help fulfil the given task [45]. Take for example visual recognition; one can use the whole image as an input instead of using the features extracted by handcrafted off-the-shelf methods such as appearance and geometric based methods. Over the last decade DNN has come

to enormous fame under Deep Learning and has become the fastest growing and most attracting research field in machine learning.

One of the very used deep learning architectures in the field of FER are deep belief networks (DBN) and convolutional neural networks (CNN); in this paper we are interested in CNN; it was first presented in 1997 [28], its structure relies basically on the alternative use of two kinds of basic layers respectively called convolutional layers (C layers) and sub-sampling layer (S layers) usually referred to as Pooling layer, and at the end a fully connected layer that leads to the output as represented in Figure 2. Further detailed information about CNN can be found in [22].

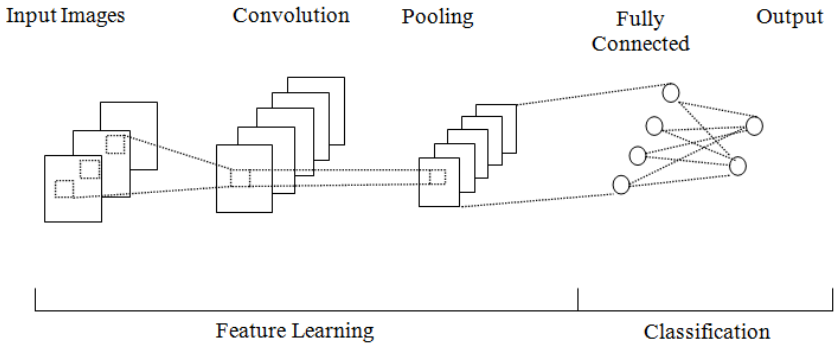

Fig. 2. Convolutional Neural Network Architecture

3 Facial Expression Recognition with Deep Learning

Over the last two decades, several approaches were developed to solve the problems of FER in computer vision, each and every one of them helped improving the recognition performance, a further survey on the elder methods can be found in [11, 44, 27, 26, 20]. An important part of the recent progress in the field was achieved thanks to the emergence of deep learning methods and specifically CNN; this section discusses the most recent and highly cited deep learning based methods that achieved high accuracy in the last five years.

A recent research presented in [5], studied the ability of CNN in improving the FER accuracy on the real image dataset of FER2013 which includes real facial images for the seven facial expressions, their solution was developed using the machine learning platform TensorFlow. the architecture consists of three convolutional layers, two max pooling layers, two fully connected layers and dropouts at different levels, authors also provided details about the different parameters of the architecture. The method achieved 91,12% in which CNN augmentation setting can provide a better performance.

Another model proposed in[52] in which face is detected from the images in datasets using Cascade Classifier, then the obtained image is transformed into grayscale level and after that normalized and at the end augmented using Image Data Generator provided by Keras API; methods used in augmentation are: horizontal flip, rotation, rescale, shear and zoom. The augmented dataset then is fed into a CNN in order to predict the emotion. Their architecture consisted of three convolutional layers with 32, 64, 128 filters respectively, kernal size is 3*3 and four fully connected layers. Data was collected from multiple databases namely: CK+, FER2013, MUG, KDEF & AKDEF and Kinface W I & II. It achieved an accuracy of 96,24% with only 120 epochs, whereas CNN without data augmentation required 260 epochs to achieve only 92,95%.

Authors in [21] proposed a single deep CNN containing convolutional layers and residual blocks. The system was trained on CK+ and JAFFE databases and achieved an accuracy of 93,24% and 95,23% respectively. It seems that combining CNN with residual blocks improves the overall results thus responds the problem of facial expression recognition and classification.

the model proposed in [37] contains three layers architecture based on deep learning in which the first two layers consist of extracting two kinds of features geometric and appearance based (LBP) then combining them using auto encoders. The third layer classifies combined features using a self-organizing map (SOM) based classifier that combines benefits of both supervised and unsupervised training algorithm; the authors consider their method as the first of a kind in this field. The system was validated on two databases CK+ and MMI and achieved respectively 98,95% and 97,55%.

Authors proposed a method based on CNN combined with specific image preprocessing steps in [34]. The different steps are: i) rotation correction, ii) cropping (eliminates the background information), iii) down-sampling (performed to reduce the image size and ensure that the location of facial component is the same in all images), iiii) intensity normalization (makes the brightness and contrast of all images the same in order to decrease the complexity of the network). The input of the pre-processing phase is a set of images with faces and their eyes location and expression id and then the output is fed to a CNN that comprises two convolutional layers, two sub-sampling layers and a fully connected layer. The system studied the impact of each pre-processing step on the accuracy rate and was tested on three widely used databases (CK+, JAFFE and BV-3DFE) and achieved the best accuracy on CK+ with 96,76%, the accuracy was improving with the addition of each pre-processing step.

Another system composed of four modules is presented in [47], the modules are: input, pre-processing, recognition and output. It was tested on two databases JAFFE an CK+ and it was compared to a k nearest neighbour (KNN) algorithm, it achieved respectively 76,7442% and 80,303%. The system was realised by implementing face detection using Haar like features in OpenCV and histogram equalization. The system architecture is composed of four layers; (two convolutional and two sub sampling) then a Softmax classifier is used for multiclassification. The authors favoured Haar like features in order to capture the

useful portion of facial expression and removing most of the background information then applied Histogram Equalization in order to make gray scale values and contrast more uniform in all images, at the end authors concluded that CNN gives better results in solving the problem of facial expression recognition after comparing their results to methods that are based on KNN.

Authors in [12] developed a CNN with variable depth giving the user the freedom to chose the number of convolutional and fully connected layers as well as the existence of batch normalization dropout and max pooling layers, also the number of filters, strides and zero padding can be specified by users with the existence of default values in case the user did not. The authors implemented the modal in Torche with high GPU capacity integration; the test and training were performed on a database provided by Kaggle Website which consists of gray scale images of faces. Their process lies on combining features extracted using the convolutional layer with those of Histogram of Oriented Gradient (HOG) and provides them as input features into fully connected layers. The system was trained under two architectures in order to decide which is the best for the recognition, the training was for 30 epochs and batch size of 128. The first shallow architecture is composed of two convolutional layers, one fully connected layer and a hidden layer composed of 512 neurons; it achieved 55% on the validation and 54% on the test set. The second one is a deeper CNN with four Convolutional layers, two fully connected layers with the first hidden layer composed of 256 neurons and the second 512; it achieved 65% on validation and 64% on test set. The authors also trained deeper architectures with five and six convolutional layers but the results were not acceptable. They also attempted to combine features from convolutional layers of both architectures with HOG features but it appears that hybrid features did not help in improving the model accuracy therefore they concluded that a not very deep architecture is fair enough to get good accuracy and solve the problem.

A real time model was presented in [2], authors proposed a method for real time vision system which detects faces and classifies them by gender and then by emotion based on CNN in which they proposed two different architectures. The first one relies on completely eliminating the fully connected layers. The second in addition to the elimination of fully connected layers, it is combined with depth wise separable convolutions and residual models. The gender and emotion classification took 0,22 ms. For gender classification the first architecture achieved 96% while the second 95% applied on IMDB gender dataset, for emotion classification both architectures were performed on FER2013 and achieved 66% accuracy.

The work presented [38] consists of a deep multi layer network for saliency prediction to get the intensity maps then pass them as an input to the CNN based AlexNet, the model was trained on the ILSVRC2012 dataset with its both posed databases CFEE and RaFD, faces were cropped using Viola and Jones algorithm, the system was trained on CFEE using 100 epochs and tested independently on RaFD and CFEE and achieved respectively an accuracy of 95,71% and 74,79%.

3D inception ResNet architecture was proposed in [18]; it focuses on extracting both spatial and temporal features from video images in video sequences. They enhanced their method by using another input which is Facial Landmarks in order to help extracting the main component of the face (eyebrows, lip corners, eyes); the model was tested by subject independent task in which every database is splited into training and validation sets and in cross validation task in which the system is trained on one database and tested on the other, the employed databases are CK+, MMI, FERA and DISFA. They achieved the following accuracies respectively in subject independent task: 93,21% , 77,50% , 77,42% and 58,00%. In addition they achieved these accuracies on cross validation task: 67,52% , 54,76% , 41,93% and 40,51%. It seems that subject independent evaluation gives better results but in both tasks authors affirmed that their method outperforms many of the state of the art methods.

DNN architecture presented in [40] was composed of two convolutional layers each followed by max pooling then inception layers. The system takes input images from seven different databases and classifies them into the six basic emotions; the databases with their accuracies for subject independent validation are: CMU MultiPIE 94,7% , MMI 77,9%, CK+ 93,2%, DISFA 55,0%, FERA76,7%, SFEW 47,7% and FER2013 66,4%. The authors observed that the use of inception layers with CNN instead of conventional CNN increases the classification accuracy on both subject independent and cross database evaluation tasks and they confirmed that results in the subject independent tests were either comparable or better than the state of the art at that time.

Authors proposed a CNN model based on structured subnets in [32], each subnet is a compact CNN model trained separately, and the final network was constructed by concatenating all the subnets together. The designed three subnets contain three, four, five Convolutional layers respectively and other parameters were identical. The general workflow of the system is composed of two stages: i) feed the input data to the subsets. ii) Predict the emotion based on the last output (stage i). The features collected by the three subsets are concatenated by adding a fully connected layer at the end, and then a Softmax layer is used as the output layer of the whole network. The best single subnet achieved 62,44% and the entire model accuracy scored 65,03%.

The work presented in [29] proposed a system that uses a webcam which can detect faces and recognize users with a distance of 2-3 m for TV environment and can recognize human emotions based on the six basic emotions; their model was developed under CAFFE and based on CUDA. It consists of three convolutional layers and two fully connected layers and rectified liner unit (ReLu) activation function. The system was trained using FER2013 and tested using real time images; it achieved a good accuracy and recognized the six primary emotions and three secondary ones (exited, bored and concentration) therefore it seems to be successful and can be used in various domains (interactive TV, intelligent vehicles and others).

An action unit inspired deep network was proposed by [33], authors were inspired by the interpretation of AU; they built a convolutional layer and max-

pooling layer and the micro action pattern representation to capture the local appearance variations caused by facial expressions; then combined the different maps and at the end employed a multi-layer learning process to generate high level features that are used for the expression recognition. authors performed cross-database validation and compared obtaind results with state of the art methods, also they compared their results with hand crafted methods and it seems that this method achieved good accuracy on three databases; CK+, MMI and SFEW.

The work presented in [30] consists of a system that recognizes emotions from static facial images using CNN. Their method involves passing pre-processed images as an input to the CNN instead of RGB images in order to minimise the network effort and complexity, then the images are converted to LBP codes to overcome the illumination changes. the model is trained on CASIA webface images and tested on the Emotion Recognition in the Wild Challenge (EmotiW 2015) and Static Facial Expression Recognition sub-Challenge (SFEW)This method at its time achieved 15,36% improvement over the baseline results and boosted the performance of the system by 40% by looking beyond RGB images as input to the CNN network.

For further works and methods in the field of image processing using convolutional neural networks the reader may refer to [56, 59, 58, 61, 63, 61, 50, 4]; these papers show the effectiveness of CNN in lots of fields such as: face recognition, gender recognition. Other works are a bit old to critisize since their results were overcomed by more recent works i.e. they are no longer interesting to investigate.

4 Comparative Study and Discussion

After studying a humble number of papers, these articles are the most recent works in domain of facial expression recognition and specifically facial expression recognition based on deep learning more specially on convolutional neural networks, we observed the following points:

Most of the presented works are based on the CNN architecture with different depth size. Even though it requires a large training dataset and a good GPU capacity but each time it is used, CNN proves its efficiency in resolving image recognition problems because of its classification capacity.

A pre-processing phase is performed almost in every work i.e. the pre-processing operations such as face detection, cropping, illumination normalization, resizing, flipping to name few, can help in improving the accuracy and decreasing the complexity of the architecture therefore enhancing the training time.

Convolotional neural networks require a huge amount of data but in case the dataset provided is not large enough, data augmentation is a key step to inlarge and expand the dataset by exercing some predefined processing such as horizontal and vertical flipping, rescaling, zooming, rotating.

Deep learning is very trendy in solving facial expression recognition problems but that doesnt mean that increasing the depth of the architecture can for sure increase the recognition rate [12], sometimes a shallow one can do the job.

Table 1. List of Presented Papers

Paper	Year	Dataset	Accuracy	Architecture
[5]	2019	FER2013	91,12%	3 C layers, 2 max P layers, 2 FC layers, dropouts at different levels
[52]	2019	Multiple Databases	96,24%	3 C layers, 4 FC layers, data augmentation
[21]	2019	CK+ JAFFE	93,24% 95,24%	C layers combined with residual blocks
[37]	2018	CK+ MMI	98,95% 97,55%	3 C layers, geometric and appearance feautures, classification is done using self-organizing map (SOM)
[34]	2017	CK+, JAFFE , BV-3DFE	96,76%	2 C layers, 2 sub-sampling layers, FC layer, pre-processing steps
[47]	2017	JAFFE CK+	76,7442% 80,303%	2 C layers, 2 sub sampling layers, Softmax classifier, Haar like features
[12]	2017	Kaggle Website	64%	CNN combined with HOG, Architecture 1 (2 C layers, 1 FC layer), Architecture 2 (4 C layers, 2 FC layers), parameters defined by the user
[2]	2017	FER2013	66%	Gender and emotion classification, Architeture 1 (eliminating FC layers), Architecture 2 (elimination combined with residual models)
[38]	2017	CFEE RaFD	95,71% 74,79%	CNN architecture based on AlexNet
[18]	2017	CK+ , MMI FERA , DISFA	93,21% ; 77,50% 77,42% ; 58,00%	CNN combined with facial landmarks
[40]	2016	CMU MultiPIE, MMI CK+, DISFA FERA, SFEW, FER2013	94,7% ; 77,9% 93,2% ; 55,0% 76,7% ; 47,7% ; 66,4%	2 C layers, max pooling layer, inception layer
[29]	2016	FER2013	-	3 C layers, 2 FC layers, rectified liner unit (ReLu) activation function
[33]	2015	CK+, MMI SFEW	93,70% ; 75,85% 30,14%	1 C layer, 1 max pooling layer, micro action pattern representation
[30]	2015	CASIA webface images EmotiW 2015, SFEW	15,36% improvement	CNN with pre-processed images

All the performed validation tasks use subject independent and cross database validation, they both give good results, but for most of the presented works it seems that the highest accuracy is obtained when performing a subject indepen-

dent validation that's because cross validation is destined for simple models with few parameters but for CNN known to have huge number of parameters performing a cross validation will be extremly time consuming and hence decrease the performance of the model.

Some papers used real time images in the test phase obtained by real time face acquisition systems in order to prove the efficiency of their system in real time application such Human Computer Interaction (HCI) evaluation and advertisement services.

most of the datasets collected by authors are either one or a combination of the fullowing databases in the presented works are the extended Cohn Kanade + (CK+) [35] which contains posed, spontaneous and smiles images. The Japanese Female Facial Expression (JAFFE) [39] which contains only posed images. The Facial Expression Recognition 2013 dataset (FER2013) [15] created using Google image search API to search for images of faces corresponding to different emotions. For further details about existing databases in the field you can check [41].

We focused on investigating works from the five last years that are based on deep learning, and according to the observed results in different articles, it seems that deep learning is the new orientation for facial and emotion expression recognition problem since different authors compared their results to those of the state-of-the-art methods.

Table 1 represents the different presented papers organized by year of publication, the table also contains the different datasets used in each paper and acuuracies achieved by each method. we observe that over the years the accuracy achieved using CNN models is increasing therefor it seems that CNN architectures and more generally depp learning fulfills the task of facial expression recognition very well.

5 Conclusion and Future Work

In this paper, we investigated the most recent and more cited works in the field of FER according to Google Scholer; these works are based on deep learning architectures and mostly on convolutional architectures. We observed that researchers are more and more interested in deep learning methods because these latter are achieving good accuracies over the last five years, therefore deep learning is considered as the new generation for solving FER problems due to its efficiency in feature extraction and classification task. At the moment we are working on developing a deep architecture based on CNN that will overcome the state of the art accuracy.

References

1. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: Application to face recognition. IEEE Transactions on Pattern Analysis & Machine Intelligence (12), 2037–2041 (2006)

2. Arriaga, O., Valdenegro-Toro, M., Plöger, P.: Real-time convolutional neural networks for emotion and gender classification. arXiv preprint arXiv:1710.07557 (2017)
3. Chakraborty, A., Konar, A., Chakraborty, U.K., Chatterjee, A.: Emotion recognition from facial expressions and its control using fuzzy logic. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans **39**(4), 726–743 (2009)
4. Chang, W.J., Schmelzer, M., Kopp, F., Hsu, C.H., Su, J.P., Chen, L.B., Chen, M.C.: A deep learning facial expression recognition based scoring system for restaurants. In: 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC). pp. 251–254. IEEE (2019)
5. Christou, N., Kanojiya, N.: Human facial expression recognition with convolution neural networks. In: Third International Congress on Information and Communication Technology. pp. 539–545. Springer (2019)
6. Cohen, I., Sebe, N., Gozman, F., Cirelo, M.C., Huang, T.S.: Learning bayesian network classifiers for facial expression recognition both labeled and unlabeled data. In: 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings. vol. 1, pp. I–I. IEEE (2003)
7. Darwin, C., Prodger, P.: The expression of the emotions in man and animals. Oxford University Press, USA (1998)
8. Dash, M., Liu, H.: Feature selection for classification. Intelligent data analysis **1**(1-4), 131–156 (1997)
9. Ekman, P., Friesen, W.V.: Constants across cultures in the face and emotion. Journal of personality and social psychology **17**(2), 124 (1971)
10. Ekman, R.: What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS). Oxford University Press, USA (1997)
11. El Ayadi, M., Kamel, M.S., Karray, F.: Survey on speech emotion recognition: Features, classification schemes, and databases. Pattern Recognition **44**(3), 572–587 (2011)
12. Eusebio, J.M.A.: Convolutional neural networks for facial expression recognition (2016)
13. Friesen, E., Ekman, P.: Facial action coding system: a technique for the measurement of facial movement. Palo Alto **3** (1978)
14. Ghimire, D., Lee, J.: Geometric feature-based facial expression recognition in image sequences using multi-class adaboost and support vector machines. Sensors **13**(6), 7714–7734 (2013)
15. Goodfellow, I.J., Erhan, D., Carrier, P.L., Courville, A., Mirza, M., Hamner, B., Cukierski, W., Tang, Y., Thaler, D., Lee, D.H., et al.: Challenges in representation learning: A report on three machine learning contests. In: International Conference on Neural Information Processing. pp. 117–124. Springer (2013)
16. Happy, S., Routray, A.: Automatic facial expression recognition using features of salient facial patches. IEEE transactions on Affective Computing **6**(1), 1–12 (2014)
17. Harms, M.B., Martin, A., Wallace, G.L.: Facial emotion recognition in autism spectrum disorders: a review of behavioral and neuroimaging studies. Neuropsychology review **20**(3), 290–322 (2010)
18. Hasani, B., Mahoor, M.H.: Facial expression recognition using enhanced deep 3d convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 30–40 (2017)

19. Heiselet, B., Serre, T., Pontil, M., Poggio, T.: Component-based face detection. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001. vol. 1, pp. I–I. IEEE (2001)
20. Jafri, R., Arabnia, H.R.: A survey of face recognition techniques. Jips **5**(2), 41–68 (2009)
21. Jain, D.K., Shamsolmoali, P., Sehdev, P.: Extended deep neural network for facial emotion recognition. Pattern Recognition Letters **120**, 69–74 (2019)
22. Karpathy, A., et al.: Cs231n convolutional neural networks for visual recognition. Neural networks **1** (2016)
23. Kobayashi, H., Hara, F.: Recognition of six basic facial expression and their strength by neural network. In: [1992] Proceedings IEEE International Workshop on Robot and Human Communication. pp. 381–386. IEEE (1992)
24. Koelstra, S., Pantic, M., Patras, I.: A dynamic texture-based approach to recognition of facial actions and their temporal models. IEEE transactions on pattern analysis and machine intelligence **32**(11), 1940–1954 (2010)
25. Kosti, R., Alvarez, J.M., Recasens, A., Lapedriza, A.: Emotion recognition in context. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1667–1675 (2017)
26. Kristensen, R.L., Tan, Z.H., Ma, Z., Guo, J.: Binary pattern flavored feature extractors for facial expression recognition: An overview. In: 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). pp. 1131–1137. IEEE (2015)
27. Kumari, J., Rajesh, R., Pooja, K.: Facial expression recognition: A survey. Procedia Computer Science **58**, 486–491 (2015)
28. Lawrence, S., Giles, C.L., Tsoi, A.C., Back, A.D.: Face recognition: A convolutional neural-network approach. IEEE transactions on neural networks **8**(1), 98–113 (1997)
29. Lee, I., Jung, H., Ahn, C.H., Seo, J., Kim, J., Kwon, O.: Real-time personalized facial expression recognition system based on deep learning. In: 2016 IEEE International Conference on Consumer Electronics (ICCE). pp. 267–268. IEEE (2016)
30. Levi, G., Hassner, T.: Emotion recognition in the wild via convolutional neural networks and mapped binary patterns. In: Proceedings of the 2015 ACM on international conference on multimodal interaction. pp. 503–510. ACM (2015)
31. Li, S.Z., Zou, X., Hu, Y., Zhang, Z., Yan, S., Peng, X., Huang, L., Zhang, H.: Real-time multi-view face detection, tracking, pose estimation, alignment, and recognition. IEEE CVPR Demo Summary (2001)
32. Liu, K., Zhang, M., Pan, Z.: Facial expression recognition with cnn ensemble. In: 2016 international conference on cyberworlds (CW). pp. 163–166. IEEE (2016)
33. Liu, M., Li, S., Shan, S., Chen, X.: Au-inspired deep networks for facial expression feature learning. Neurocomputing **159**, 126–136 (2015)
34. 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 Recognition **61**, 610–628 (2017)
35. Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z., Matthews, I.: The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops. pp. 94–101. IEEE (2010)
36. Lucey, P., Cohn, J.F., Matthews, I., Lucey, S., Sridharan, S., Howlett, J., Prkachin, K.M.: Automatically detecting pain in video through facial action units. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) **41**(3), 664–674 (2010)

37. Majumder, A., Behera, L., Subramanian, V.K.: Automatic facial expression recognition system using deep network-based data fusion. IEEE transactions on cybernetics **48**(1), 103–114 (2016)
38. Mavani, V., Raman, S., Miyapuram, K.P.: Facial expression recognition using visual saliency and deep learning. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2783–2788 (2017)
39. Michael, J., Lyons, M.K., Gyoba, J.: Japanese female facial expressions (jaffe). Database of digital images (1997)
40. Mollahosseini, A., Chan, D., Mahoor, M.H.: Going deeper in facial expression recognition using deep neural networks. In: 2016 IEEE Winter conference on applications of computer vision (WACV). pp. 1–10. IEEE (2016)
41. Mollahosseini, A., Hasani, B., Mahoor, M.H.: Affectnet: A database for facial expression, valence, and arousal computing in the wild. IEEE Transactions on Affective Computing **10**(1), 18–31 (2017)
42. Pentland, A., Moghaddam, B., Starner, T., et al.: View-based and modular eigenspaces for face recognition (1994)
43. Rowley, H.A., Baluja, S., Kanade, T.: Neural network-based face detection. IEEE Transactions on pattern analysis and machine intelligence **20**(1), 23–38 (1998)
44. Sandbach, G., Zafeiriou, S., Pantic, M., Yin, L.: Static and dynamic 3d facial expression recognition: A comprehensive survey. Image and Vision Computing **30**(10), 683–697 (2012)
45. Schmidhuber, J.: Deep learning in neural networks: An overview. Neural networks **61**, 85–117 (2015)
46. Schneiderman, H., Kanade, T.: A statistical approach to 3D object detection applied to faces and cars. Carnegie Mellon University, the Robotics Institute (2000)
47. Shan, K., Guo, J., You, W., Lu, D., Bie, R.: Automatic facial expression recognition based on a deep convolutional-neural-network structure. In: 2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA). pp. 123–128. IEEE (2017)
48. Sprengelmeyer, R., Young, A., Mahn, K., Schroeder, U., Woitalla, D., Büttner, T., Kuhn, W., Przuntek, H.: Facial expression recognition in people with medicated and unmedicated parkinsons disease. Neuropsychologia **41**(8), 1047–1057 (2003)
49. Sung, K.K., Poggio, T.: Example-based learning for view-based human face detection. IEEE Transactions on pattern analysis and machine intelligence **20**(1), 39–51 (1998)
50. Tang, J., Zhou, X., Zheng, J.: Design of intelligent classroom facial recognition based on deep learning. In: Journal of Physics: Conference Series. vol. 1168, p. 022043. IOP Publishing (2019)
51. Tian, Y., Kanade, T., Cohn, J.F.: Facial expression recognition. In: Handbook of face recognition, pp. 487–519. Springer (2011)
52. Uddin Ahmed, T., Hossain, S., Hossain, M.S., Ul Islam, R., Andersson, K.: Facial expression recognition using convolutional neural network with data augmentation. In: Joint 2019 8th International Conference on Informatics, Electronics & Vision (ICIEV) (2019)
53. Valstar, M.F., Almaev, T., Girard, J.M., McKeown, G., Mehu, M., Yin, L., Pantic, M., Cohn, J.F.: Fera 2015-second facial expression recognition and analysis challenge. In: 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG). vol. 6, pp. 1–8. IEEE (2015)
54. Valstar, M.F., Jiang, B., Mehu, M., Pantic, M., Scherer, K.: The first facial expression recognition and analysis challenge. In: Face and Gesture 2011. pp. 921–926. IEEE (2011)

55. Viola, P., Jones, M.J.: Robust real-time face detection. International journal of computer vision **57**(2), 137–154 (2004)
56. Wu, Y., Hassner, T., Kim, K., Medioni, G., Natarajan, P.: Facial landmark detection with tweaked convolutional neural networks. IEEE transactions on pattern analysis and machine intelligence **40**(12), 3067–3074 (2017)
57. Yang, J., Zhang, D., Frangi, A.F., Yang, J.y.: Two-dimensional pca: a new approach to appearance-based face representation and recognition. IEEE transactions on pattern analysis and machine intelligence **26**(1), 131–137 (2004)
58. Yu, Z., Zhang, C.: Image based static facial expression recognition with multiple deep network learning. In: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction. pp. 435–442. ACM (2015)
59. Zeng, N., Zhang, H., Song, B., Liu, W., Li, Y., Dobaie, A.M.: Facial expression recognition via learning deep sparse autoencoders. Neurocomputing **273**, 643–649 (2018)
60. Zhang, L., Tjondronegoro, D.: Facial expression recognition using facial movement features. IEEE Transactions on Affective Computing **2**(4), 219–229 (2011)
61. Zhang, T., Zheng, W., Cui, Z., Zong, Y., Yan, J., Yan, K.: A deep neural network-driven feature learning method for multi-view facial expression recognition. IEEE Transactions on Multimedia **18**(12), 2528–2536 (2016)
62. 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 **4**, 8375–8385 (2016)
63. Zhao, X., Shi, X., Zhang, S.: Facial expression recognition via deep learning. IETE technical review **32**(5), 347–355 (2015)
64. Zhao, X., Zhang, S.: A review on facial expression recognition: Feature extraction and classification. IETE Technical Review **33**(5), 505–517 (2016)