

A Review on Facial Emotion Recognition Using Machine and Deep Learning Algorithm

Dr. Ramesh Y. Mali, Department of E&C, MITSOE, Pune, India, ramesh_malius@yahoo.com

Prof. Umashankar More, Department of E&C, MITSOE, Pune, India, more@mituniversity.edu.in

Abstract Facial emotions are the changes in facial expressions about a person's inner excited tempers, objectives, or social exchanges which are scrutinized with the aid of computer structures that attempt to subsequently inspect and identify the facial feature and movement variations from visual data. Facial emotion recognition (FER) is a noteworthy area in the arena of computer vision and artificial intelligence due to its significant commercial and academic potential. FER has become a widespread concept of deep learning and offers more fields for application in our day-to-day life. Facial expression recognition (FER) has gathered widespread consideration recently as facial expressions are thought of as the fastest medium for communicating any of any sort of information. Recognizing facial expressions provides an improved understanding of a person's thoughts or views. With the latest improvement in computer vision and machine learning, it is plausible to identify emotions from images. Analyzing them with the presently emerging deep learning methods enhance the accuracy rate tremendously as compared to the traditional contemporary systems. This paper emphasizes the review of a few of the machine learning, deep learning, and transfer learning techniques used by several researchers that flagged the means to advance the classification accurateness of the FEM.

Keywords — *Facial expression recognition, deep learning, machine learning, transfer learning*

I. INTRODUCTION

When it comes to social communication facial language is considered to be enormously crucial. Usually, communication encompasses nonverbal as well as verbal communication. Facial expressions are a tool of non-verbal communication. Large communication can be conveyed by delicately through facial expressions. Non-verbal communication means communicating through body language, gesture, facial expressions, eye contact, and paralanguage. The facial expression of human emotion can generate both everyday and technical application past laboratory experiments. Identifying human emotions has garnered the researchers' attention, as the competence of recognizing expressions aids in human-computer interaction by improving human emotional intelligence ("EQ"). Facial Recognition is the technology that involves techniques and methods to detect the emotions from facial expression. Various technological advancements in the area of Artificial Intelligence and Machine Learning, have made the emotion recognition simpler. In traditional FER methods, the FER comprises three main steps,

- face and facial element recognition,
- feature extraction, and
- classifying expression.

In the primary step, an input image is used to detected a face image from it, and facial constituents such as eyes and nose or other landmarks are also identified from the facial region.

For the second step, several temporal and spatial features are taken out from the facial constituents. There exist two kinds of features that are usually used in representing facial appearance: appearance and geometric features.

Appearance features signify the alteration in texture of face when a specific action is executed whereas geometric features gauge the displacements of specific parts of the face such as mouth corners or brows.

The third step is a classification which is one of the significant procedures in which emotions such as sad, smile, disgust, anger, fear, and surprise are classified. In this stage previously trained FE classifiers, for example, an Adaboost, random forest, and support vector machine (SVM), generate the recognition outcomes with the help of the previously extracted features.

Generally, there are seven kinds of emotions viz. Anger, Disgust, Fear, Happy, Neutral, Sad, and Surprise. Fig. 1. shows all the seven emotions visually.



(a)

(b)

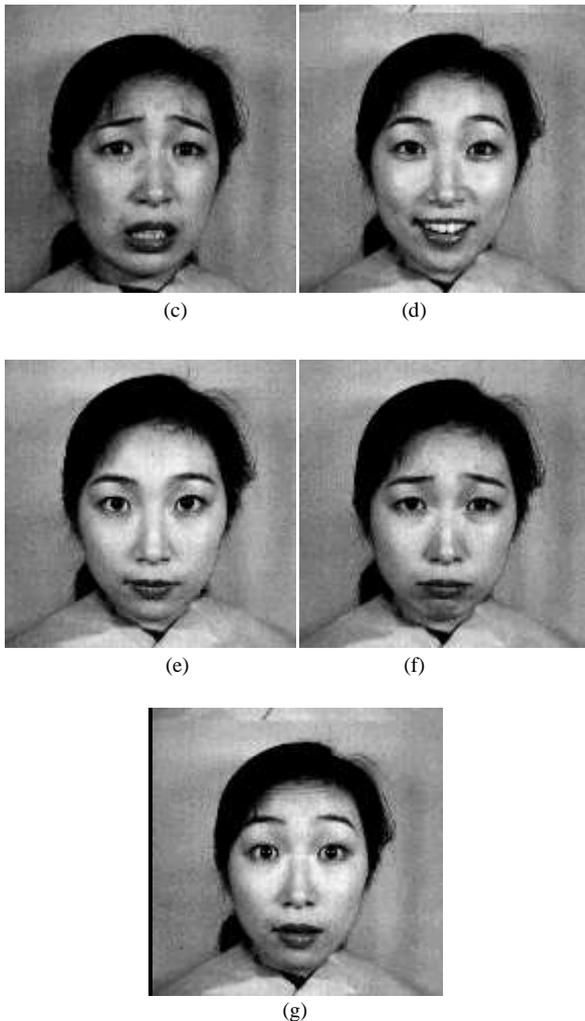


Fig. 1. Basic facial expression (a)Anger, (b)Disgust, (c)Fear, (d)Happy, (e)Neutral, (f)Sad, and (g)Surprise

The positions of the eyes, mouth, eyebrows according to the emotions are defined in Table I

TABLE I
Facial Expression Description Of Seven Basic Emotions

Sr. No.	Emotions Class	Description of facial expression			
		Eyes	Eyebrows	Lips	Mouth
1	Anger	widely opened	downward	closed tightly.	Relaxed
2	Disgust	Relaxed	Relaxed eyebrows	the upper lip is raised and curled	Relaxed
3	Fear	open and tense.	raised and pulled together.	Relaxed	Relaxed
4	Happy	Relaxed	Relaxed.	Relaxed	Open and corners upturned.
5	Neutral	Relaxed	Relaxed	Relaxed	Relaxed
6	Sad	slightly closed	Inner eyebrows are bent upward.	Relaxed	Relaxed
7	Surprise	Widely Opened	raised	the upper and lower lip is raised	opened

II. MACHINE LEARNING APPROACH FOR FACIAL EXPRESSION RECOGNITION

The generalized block diagram of the machine learning approach for FER is shown in Fig.2.

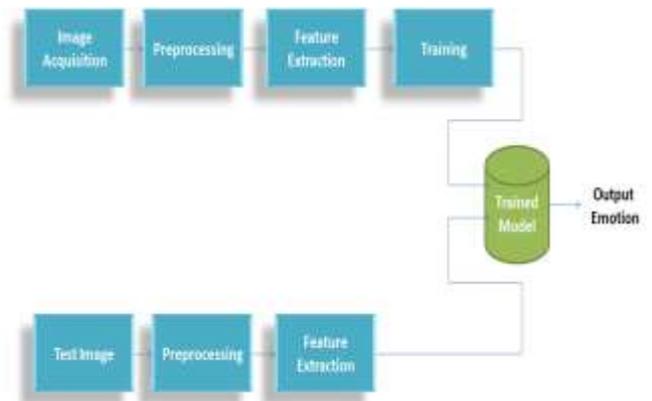


Fig. 2. Basic block diagram of FER system using a Machine learning algorithm

SVM is a machine learning technique that practices four kinds of kernels for its improved performance. It involves two kinds of approaches. They are one against all and one against one approach. One against all classification suggests that it builds one trial for every class[1]. One against one classification suggests that it builds one class for every duo of classes [2].

SVM is a robust classification approach for enhanced dimensionality difficulties.SVM classifier has better utility as compared with the rest of the classifiers for facial expression and emotion recognition. Comparing various classifiers SVM classifier provides recognition with enhanced precision and also gives better classification. Therefore SVM classifier is an exceedingly utilized classifier for FER and it employs a one-to-one, one-to-all grouping method. The following are some of the researches done previously in the FER field using machine learning.

The different approach for facial expression recognition using machine learning approach is discussed below.

Ghimire and Lee [3] made use of two types of geometric features that gave the association among facial components which help to build a feature vector that is further used for training. For the classifiers, they have presented two methods, as using an SVM on the advanced feature vectors or using multi-class Adaboost with the active time warping. The features of appearance are generally taken out from the comprehensive facial region [4] or diverse regions of the face that encompass various kinds of information [5,6].

The example of using global features is seen in Happy et al. [4] which used a Local Binary Pattern (LBP) histogram of dissimilar block sizes from a comprehensive facial part as the feature and segregated numerous facial expressions using principal component analysis (PCA). Albeit this method is executed in real-time, the recognition precision

seems to be reduced because it is not able to reflect local variation present in the facial components into the feature vector.

Ghimire [7] extracted area-specific features of appearance by distributing the complete face region into area-specific local regions. Significant local regions are recognized using an incremental search technique, that gave reduced feature dimensions, and improved recognition accuracy.

III. DEEP LEARNING APPROACH FOR FACIAL EXPRESSION RECOGNITION

The generalized block diagram of the deep learning approach for FER is presented in Fig.3.

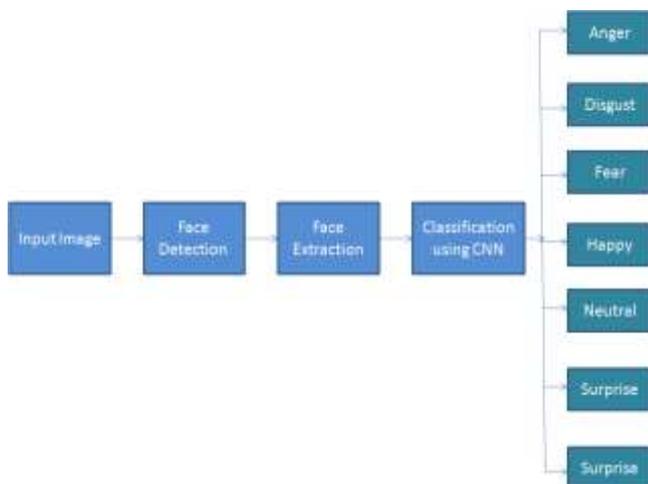


Fig. 3. Basic block diagram of the facial emotion recognition system using a deep learning algorithm

Deep-learning-based FER approaches drastically reduced the reliance on face-physics-centered procedures and other pre-processing methods by empowering “end-to-end” learning to occur directly from the input images [8]. Among the various deep-learning techniques available, the convolutional neural network (CNN), is the utmost popular network model. In CNN-centered techniques, a filter collection is present in the convolution layers where the input images get convolved to generate a feature map. After that, every feature map is joined to fully connected networks, and the face emotion is detected. Convolution Neural Network (CNN) comprises of two layers that are convolutional layer and sub-sampling layer where the two-dimensional images are used as input. In the convolutional layer, the feature maps are generated with the help of intricate convolution kernels [9].

In FER, the CNN classifier is employed as manifold classifiers for the various areas of the face. For example, if CNN is outlined for a complete image of a face, after that the initial frame for CNN is taken from the area of the mouth and then for the area of eye in this way for every area CNNs are outlined.

In recent times, there have been advancements in deep-

learning algorithms that are employed in the area of computer vision, which includes a recurrent neural network (RNN). These deep-learning-centered algorithms are exploited for extracting features, classifying, and recognizing jobs. CNN has three kinds of heterogeneous layers: convolution layer, max-pooling layer, and fully connected layer.

Emotion is identified from facial images using filter banks and Deep CNN [10] which gives a high accuracy rate with which we had an inference that deep learning can also be employed for emotion detection. Facial emotion recognition can be also performed using image spectrograms with deep convolutional networks that are implemented in [11].

This paper [12] studies different databases used for facial emotion recognition, features selected from facial expression images, classifiers used to classify different classes of emotions. As the amount of data array is taken and the method of a bottleneck is used, Long Short-term Memory (LSTM) is used for Facial emotion recognition [13]. Emotion has been recognized from facial expressions using hidden Markov models and deep belief networks with unweighted average recall (UAR) of about 56.36% (~) [14].

Numerous algorithms are presented to understand the deep learning techniques to produce models that can classify with accuracy.

IV. DEEP LEARNING APPROACH WITH TRANSFER LEARNING FOR FACIAL EXPRESSION RECOGNITION

Transfer learning is another type of deep learning methods which uses the knowledge obtained from solving a problem to solve another problem. Transfer learning indeed solves problems within a short duration of time. Transfer Learning is implemented whenever the computation cost has to be reduced and to achieve accuracy with less training. Transfer learning is a regularly applied technique that uses learned weights of a model (e.g. ImageNet) and implements those by fixing other layers and holding the rest of the layers in the network.

The modern application of transfer learning for emotion detection can be classified into two key methods. First mentions the accessibility of labels in the target area. Supervised transfer learning comprises information from labeled data for target as well as source area while the learning takes place, while unsupervised learning has information only from the label of the source area [15]. Unsupervised transfer learning allows the building of trustworthy machine learning methods, in those areas where labeled data are not available. The second classification denotes to the accessibility of one or several databases in the source area. Single source transfer learning consists of a single database while multisource transfer learning consists of multiple sets of data [16].

Real-time facial emotion recognition is done through RGB image classification using transfer learning methodologies in which knowledge gained from solving a problem and that is implemented for another problem [17].

The novel approach of transfer learning includes obtaining high-level features with the help of a CNN trained on huge sources of data [18,19,20] or transferring the knowledge from greater quality auxiliary image datasets.

V. DISCUSSION

Conventional approaches to machine learning need comparatively lower memory and computing power than deep learning-driven approaches. Hence, such approaches are still under study to be used in real-time systems for their lesser computational intricacy and a higher degree of precision [21]. However, feature extraction and the classifiers are required to be programmer designed in machine learning approaches and they cannot be mutually optimized to enhance the system performance [22,23].

As against this, deep learning-based approaches suggest classifiers and features by deep neural network professionals. Deep learning-based approaches extract optimum features that have the preferred characteristics straight away from data with the aid of deep convolutional neural networks. Though, it is not very easy to gather a huge amount of training data for the facial emotion under the different conditions that are enough deep neural networks to learn. Besides, deep learning-based approaches entail more of a higher-level and massive computing device than machine learning approaches to conduct testing and training [24]. Hence, it is essential to decrease the computational load at the conclusion time of deep learning algorithms. Speaking of Image-based transfer learning, it is mostly employed to dig out high-level features present within images with the help of CNN. By utilizing the latest transfer learning methods, or by applying additional kinds of signals it is possible to enhance transfer learning and improve emotion recognition accuracy. Discovering the ability to transfer knowledge from lab collated data to data gathered from real-life situations can considerably enhance the capacity of emotion recognition in real-time applications.

VI. CONCLUSION

Emotion recognition is a noteworthy field of research in the existing situation. Emotion identification executed with the aid of deep learning techniques is the flourishing practice used by scholars to create inventive models. The analysis of emotions is a challenging work due to the inherent uncertainty in the interpretation of hidden mental states from behavioral signs. As the facial expression recognition systems are becoming robust and effective in communications, many other innovative applications and uses are yet to be seen. This research paper targets to deliver a brief outline of the facial emotion recognition

system developed with the help of machine learning, deep learning, and transfer learning algorithms. The review of different machine learning, deep learning, and transfer learning algorithms are stated to categorize the emotions.

After reviewing different approaches on the facial expression it is observed that initially facial landmarks were used to recognize facial expression. The position of nose, eyes, eyebrows, mouth, and lip was considered for recognition. Later on, feature extraction techniques like PCA, LDA, etc. were used to describe the facial expression and machine learning algorithms were used to classify the seven emotions. Recently, with the advancement in the processing speed of the hardware platform, deep learning algorithms were used to improve the accuracy and generalization of the model. With the deep learning models, transfer learning approaches helps to improve the accuracy of the system.

REFERENCES

- [1] G. Zhao, M. Pietikainen, "Boosted multi-resolution spatiotemporal descriptors for facial expression recognition", *Pattern Recognition. Lett.* 30, pp.1117–1127.
- [2] S. Biswas, "An Efficient Expression Recognition Method using Contourlet Transform", *Int. Conf. Percept. Mach. Intell.* pp. 167–174.
- [3] D. Ghimire, J. Lee, "Geometric feature-based facial expression recognition in image sequences using multi-class AdaBoost and support vector machines", *Sensors* 2013, 13, pp. 7714–7734.
- [4] S. L. Happy, A. George, A. Routray, "A real-time facial expression classification system using local binary patterns", In *Proceedings of the 4th International Conference on Intelligent Human-Computer Interaction*, Kharagpur, India, 2012, pp. 1–5.
- [5] M. H. Siddiqi, R. Ali, A. M. Khan, Y. T. Park, S. Lee, "Human facial expression recognition using stepwise linear discriminant analysis and hidden conditional random fields", *IEEE Trans. Image Proc.* 2015, 24, pp. 1386–1398.
- [6] R. A. Khan, A. Meyer, H. Konik, S. Bouakaz, "Framework for reliable, real-time facial expression recognition for low-resolution images", *Pattern Recognit. Lett.* 2013, 34, pp. 1159–1168.
- [7] D. Ghimire, S. Jeong, J. Lee, S. H. Park, "Facial expression recognition based on local region specific features and support vector machines", *Multimed. Tools Appl.* 2017, 76, pp. 7803–7821.
- [8] K. Shan, J. Guo, W. You, D. Lu, R. Bie, "Automatic Facial Expression Recognition Based on a Deep Convolutional-Neural Network Structure", *IEEE 15th Int. Conf. Softw. Eng. Res. Manag. Appl.* pp.123–128.

- [9] R. Walecki, O. Rudovic, V. Pavlovic, B. Schuller, M. Pantic, "Deep structured learning for facial action unit intensity estimation", In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017, pp. 3405–3414
- [10] K. Y. Hueng, C. H. Wiu, T. H. Yieng, M. H. Sha, and J. H. Chiu, "Emotion recognition using autoencoder bottleneck features and LSTM," in 2016 International Conference on Orange Technologies (ICOT), 2016, pp. 1–4.
- [11] M. N. Sttilar, M. Leich, R. S. Bolie, and M. Skinter, "Real-time emotion recognition using RGB image classification and transfer learning," 2017 11th Int. Conf. Signal Process. Commun. Syst., pp. 1–8, 2017.
- [12] D. Le and E. M. Provost, "Emotion recognition from spontaneous speech using Hidden Markov models with deep belief networks," 2013 IEEE Workshop on Automatic Speech Recognition and Understanding, Olomouc, 2013, pp. 216–221.
- [13] Jiang Haihua, HuBinLiu Zhenyu, Yan Lihua, Wang Tianyang, Liu Fei, Kang Huanyu, Li Xiaoyu, "Investigation of emotions for identifying depression using various classifiers," Expression Commun., vol. 90, pp. 39–46, 2017.
- [14] P. Hairar, R. Burgeet, and M. K. Duitta, "Facial Emotion Recognition with Deep Learning," in Signal Processing and Integrated Networks (SPIN), 2017, pp. 4–7.
- [15] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," in IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
- [16] Z. Ding, H. Zhao, and Y. Fu, "Multi-source Transfer Learning", Cham: Springer International Publishing.
- [17] Q. Maio, M. Diong, Z. Huaeng, and Y. Zhaen, "Learning Salient Features for Emotion Recognition Using Neural Networks," IEEE Trans. Multimed., vol. 16, no. 8, 2014.
- [18] T. Q Ngo, and S. Yoon, "Facial expression recognition on static images," in International Conference on Future Data and Security Engineering, Springer, 640–647.
- [19] H. Kaya, F. Gurpinar, and A. A. Salah, "Video-based emotion recognition in the wild using deep transfer learning and score fusion", Image Vis. Comput. 65, pp. 66–75.
- [20] S. Aly, A. Trubanova, L. Abbott, S. White, and A. Youssef, "Vtkfer: a Kinect-based rgb+d time dataset for spontaneous and non-spontaneous facial expression recognition," in 2015 International Conference on Biometrics (ICB), pp. 90–97.
- [21] M. Suk, B. Prabhakaran, "Real-time mobile facial expression recognition system - A case study", In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, USA, 24–27 June 2014, pp.132–137.
- [22] Y. Tian, P. Luo, X. Luo, X. Wang, X. Tang, "Pedestrian detection aided by deep learning semantic tasks", In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 8–10 June 2015; pp. 5079–5093.
- [23] S. Deshmukh, M. Patwardhan, A. Mahajan, "Survey on real-time facial expression recognition techniques", IET Biom.,2016, pp.155–163.
- [24] W. Wei, Q. Jia, G. Chen, "Real-time facial expression recognition for affective computing based on Kinect", In Proceedings of the IEEE 11th Conference on Industrial Electronics and Applications, Hefei, China, 5–7 June 2016, pp. 161–165.