FOREX Publication
Open Access | Rapid and quality publishing

International Journal of
Electrical and Electronics Research (IJEER)
Research Article | Volume 10, Issue 1 | Pages 18-22 | e-ISSN: 2347-470X

A New Hybrid Approach for Efficient Emotion Recognition using Deep Learning

Mayur Rahul[1], Namita Tiwari[2], Rati Shukla[3], Devvrat Tyagi[4], Vikash Yadav[5]

[1]AP, DoCA, UIET, CSJM Univ., Kanpur, UP, [2]AP, DoM, SoS, CSJM Univ., Kanpur, UP, [3]MNNIT, Prayagraj, Allahabad, UP, [4]AP, ABES Engg. College, Ghaziabad, UP, [5]Lec., DoTE, UP, India

***Correspondence:** Vikash Yadav; Email: vikas.yadav.cs@gmail.com

ABSTRACT- Facial emotion recognition has been very popular area for researchers in last few decades and it is found to be very challenging and complex task due to large intra-class changes. Existing frameworks for this type of problem depends mostly on techniques like Gabor filters, principle component analysis (PCA), and independent component analysis(ICA) followed by some classification techniques trained by given videos and images. Most of these frameworks works significantly well image database acquired in limited conditions but not perform well with the dynamic images having varying faces and images. In the past years, various researches have been introduced framework for facial emotion recognition using deep learning methods. Although they work well, but there is always some gap found in their research. In this research, we introduced hybrid approach based on RNN and CNN which are able to retrieve some important parts in the given database and able to achieve very good results on the given database like EMOTIC, FER-13 and FERG. We are also able to show that our hybrid framework is able to accomplish promising accuracies with these datasets.

General Terms: Human-computer interaction, Emotion recognition, Facial images

Keywords: Recurrent neural networks, Convolutional neural networks, Classification methods, PCA, ICA, EMOTICA, FER-13

ARTICLE INFORMATION

Author(s): Mayur Rahul, Namita Tiwari, Rati Shukla, Devvrat Tyagi, Vikash Yadav
Received: Feb 20, 2022; **Accepted**: Mar 08, 2022; **Published**: Mar 30, 2022;
e-ISSN: 2347-470X;
Paper Id: IJEER220220;
Citation: doi.org/10.37391/IJEER.100103
Webpage-link:
www.ijeer.forexjournal.co.in/archive/volume-10/ijeer-100103.html

1. INTRODUCTION

The facial emotions are the unavoidable region of communication among human beings. They can be used in various forms that cannot be easily perceived by normal eyes. That is why, using given tools, any sign following and preceding them can be subject to recognition and detection. There has been found to be increase in the requirement for the identification of human's emotions in the past decades to grow interest in the human facial emotion recognition in different fields like medicine [1], animation [2], security [3], human-computer interaction [4,5], and also the diagnosis of autism disorders in urban sound perception [6] and children [7].

The facial emotion recognition could be processed adopting various characteristics like facial expressions [8], EEG [9], and text [10]. The facial expressions are extremely popular in these features because they contain various features and are visible for efficient emotion recognition. Also, the collections of faces are so easy [11].

In the earlier years, with the help of deep learning, recognition results are significantly improved [12]. Various important features have been extracted to get facial emotion recognition system [13,14]. However, facial emotion recognition systems are only depends on the certain face regions like eyes, nose and lips, and other regions like hair and forehead doesn't take much part in the identification of emotions [15]. Therefore, we can say that most of our facial emotion recognition systems depend only on the certain part not in the other regions.

In this research, we introduced the hybrid method for efficient recognition of emotions. This method consists of two deep learning methods like CNN and RNN are used as a feature extraction technique and SVM is used as a classification technique. The main findings of our research are:

(1) We introduced hybrid method for efficient facial emotion recognition.
(2) We used combination of RNN and CNN for feature extraction and SVM for classification.
(3) We used the publicly available datasets like EMOTIC, FER-13, and FERG in our research.
(4) We are able to prove our facial emotion recognition system better than the existing systems.
(5) We also able to compare our results with all the given datasets.

The remaining part of our paper is structured as follows: similar works have been discussed in *section 2*. The presented methodology has been described in *section 3*, assessments and outcomes have been summarized in *section 4* and finally concluded in *section 5*.

2. RELATED WORKS

The first major contribution in facial expression recognition was given by Paul Ekman [16]. His framework was able to identify six basic facial expressions like surprise, fear, joy,

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 10, Issue 1 | Pages 18-22 | e-ISSN: 2347-470X

sadness, disgust, anger. Later, his framework based on Facial Action Coding System (FACS) was also able to give benchmark in this area [17]. Neutral expression was also incorporated in many datasets gives seven facial expressions.

The previous fact-finding on facial emotion recognition mostly focuses on 2-step traditional approach using machine learning [34]. The first step comprise of important feature extraction using Gabor filters, LBP, LMSP, Zernike moments etch while second step comprise of classification step using random forest, SVM, KNN is used to identify the emotions in the image[35]. These techniques are limited to small datasets while with the addition of long datasets, they unable to perform well. These problems and challenges found in these new images that having sunglasses, partial faces, dynamic background, and occlusions.

The great success of deep learning especially using CNN for the efficient classification of images and some computer vision problem, various researchers group are using deep learning concept to recognize facial emotions [18]. Khorrami et al. introduced the CNN based model to get the better accuracy in recognizing facial emotions with the help of zero biased CNN on the Toronto face dataset (TFD) and extended Cohn-kanade dataset (CK+) [19].

Aneja et al. introduced a framework for the facial emotion recognition based on deep learning using animated characters. They trained the network model of human faces. They were also able to trained animated faces and other to trained human faces with respect to animated faces [2]. Mollahosseini et al. proposed a framework based on neural networks for facial emotion recognition using one pooling layer, four inception layers and two convolutional layers [8]. Liu et al. introduced a hybrid system to combine both classification and feature extraction in one looped web accessing two parts for getting feedback. The author has used boosted deep belief network (BDBN) on JAFFE and CK+, and get the best in latest accuracy [20].

Barsoum et al. proposed a framework based on deep learning from acquisition of noisy labels using crowd-sourcing in truth images [21]. Author opt 10 taggers to rename every image in given dataset and applied different cost procedures for DCNN, achieved best result. Han et al. introduced an incremental boosting CNN called IB-CNN, for the increase in accuracy rate for the spontaneous images datasets by increasing discriminative neurons [22]. This method showed the best results at that time. Meng et al. introduced an identity-aware CNN (IA-CNN) based on identity and emotion-sensitive methods to minimize changes in identity and emotion-based information [23].

Fernandez et al. introduced end to end web framework for emotion recognition using attention based model [24]. Want et al. introduced a framework based on self-cure based network which handles uncertainty efficiently and prevents from uncertain facial emotion images [25]. Further, self-cure network put down the uncertainty from both origins: (1) a self-calculating technique over a small batch for every training sample using regularization of ranking (2) a relabeling technique to update labels of given sample in smallest ranking class. Wang et al. introduced an approach for facial emotion recognition which is efficient in real-world pic and occlusion change. They are able to introduced Region Attention Network (RAN) to importantly acquire the special features of the face region and occlusion in FER. Some of latest research found in facial emotion recognition based multiple attention networks for FER [26], deep learning based self-attention network for FER [27] and latest literature review on FER [28].

All of the discussed research achieved good accuracy over state-of-the-art works facial emotion recognition, but their technique is lack of recognizing special facial emotion recognition for expression detection. In this research, we are going to focus this drawback by introducing a system based on hybridization of RNN and CNN that are used to focus in important features and SVM is used to classify the facial emotions.

3. PROPOSED METHODOLOGY

We have proposed a system based on hybrid technique to recognize the emotions in facial image datasets. The improvement in many hybrid based systems depends on the neurons addition and adding more smooth flow in the networks. They are applicable to the classification of large number of datasets available in the real world. In the area of facial emotion recognition, we are able to show that small layers are capable of work well even in the given small datasets. We have also compared the results with the existing results using different publicly available datasets.

The facial images don't have all the regions importantly useful for the efficient recognition of facial emotions, and in the most of the cases we simply focus on the particular region to get the relevant sense to basic emotion. To overcome this problem, we proposed a system that works on combination of CNN and RNN to get the selected facial regions from the given datasets.

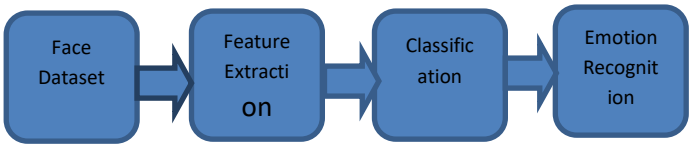

Figure 1: Facial Emotion Recognition System

The *figure 1* shows the framework used for introduced system. It basically contains four steps. The first step is to acquire the image from the face datasets. Further, feature extraction step is used to extract the important feature from the image using CNN and RNN. The feature extraction step consists of six layers, with every three following rectified activation procedure and max-pooling layer. They are then following connected layers and dropout layer. The given localization web consists of three convolution layer following pooling layer and a unit and three fully connected layers. The localization network mainly focuses on the important part of

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 10, Issue 1 | Pages 18-22 | e-ISSN: 2347-470X

the facial regions. We used affine transformation for the transition between inputs to output.

The output found from the CNN will be the output for the RNN. The LSTM (Long Short Term Memory) is the kind of RNN that have ability to transform set of input into set of output. We use the LSTM as used by the Donahue et al. [30]. After feature extraction, classification has been done using SVM. The accuracy of SVM for classifying facial images is significantly good.

4. EXPERIMENTS AND RESULTS

We are now able to produce some results using publicly available EMOTIC (18,313 images with 23,788 annotated people) [31], FER-13 (FER2013 consists of approx. 30,000 facial images of distinct expressions with size of 48×48, and the main class of it can be further subdivided into seven types: Zero=Angry, one=Disgust, two=Fear, three=Happy, four=Sad, five=Surprise, six=Neutral. The Disgust facial expression in the dataset has minimum number of images – 600, while other classes have around 5,000 samples for each class.) [32], and FERG datasets (FERG is a dataset of cartoon characters containing 55,769 annotated facial images of 6 characters. Every character are categories into 7 types of cardinal emotions, viz. surprise, sadness, neutral, joy, anger, disgust, and fear) [33]. For each case, we are able to train the given model using subpart of dataset, validated on the given validation set and accuracy calculated on test set.

Table 1: Confusion Matrix using EMOTIC dataset

	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
Angry	71.3	7.1	1.7	16.3	0.9	1.5	1.2
Disgust	9.1	75.2	6.4	2.3	2.5	3.7	0.8
Fear	9.8	2.7	76.3	1.3	1.3	4.9	3.7
Happy	3.9	6.2	1.5	64.2	4.4	3.4	2.9
Neutral	2.4	3.7	8.9	6.4	67.4	5.9	5.3
Sad	3.6	4.3	1.5	1.1	3.7	80.4	5.4
Surprise	2.7	3.9	3.5	3.9	5.4	6.9	73.7

The performance analysis has been explained on various datasets in the given section after describing the technique of our training process. We have trained the model in each and every datasets but the variables and parameters are identical in these models. We have initialized given weights using some Gaussian variables with standard deviation of 0.07. We also used L2 regularization technique with the given decay value of 0.0018. It took basically 3- 5 hours to train our model.

Table 2: Confusion Matrix using FER-13 dataset

	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
Angry	91.4	0	0	0	5.6	3.0	0
Disgust	0	100	0	0	0	0	0
Fear	0	0	93.4	3.1	0	0	3.5
Happy	0	0	1.8	98.2	0	0	0
Neutral	0	0	0	3.2	86.4	10.4	0
Sad	2.7	0	0	0	4.8	92.5	0
Surprise	0	0	0	0	3.3	0	96.7

The EMOTIC and FER-13 datasets have equal number of images while FERG contains more images. We used oversampling in order to overcome this imbalance. The data augmentation method is used to train the model on the given larger dataset i.e. FERG.

Table 3: Confusion Matrix using FERG dataset

	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
Angry	68.4	6.9	1.6	16.6	1.1	1.7	3.7
Disgust	10.3	70.2	7.8	2.4	2.9	3.9	2.5
Fear	10.3	2.9	69.8	1.9	1.7	5.1	8.3
Happy	3.6	6.6	16	60.3	4.8	3.9	4.8
Neutral	2.7	3.6	9.2	6.7	63.2	5.6	9.0
Sad	3.9	4.7	1.8	1.4	3.9	75.6	8.7
Surprise	2.6	4.6	3.9	4.3	5.9	9.5	69.2

The experiments have been conducted on the above datasets to present the performance of our model. For every dataset, we divide the entire dataset in train set, test set and validation set. The three datasets are divided as 80% for train set, 10% for test set and 10% for validation set. We trained model for each datasets in our experiments, but we have maintain all the parameters and same in all the datasets.

Table 4: Comparison of overall accuracy

Overall Accuracy	
EMOTIC [31]	72.64
FER-13 [32]	94.08
FERG [33]	68.10

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 10, Issue 1 | Pages 18-22 | e-ISSN: 2347-470X

We initialized some of its parameters as: Gaussian random parameters with SD as .0065 and mean as 0, Alternating Direction Method of Multipliers (ADMM) as .025 and L2 regularization as .0015. The average time to take training process is around 1.5 - 2 hours. The performance of our model in all datasets is depicted in *table 1, 2, 3*. The performance of our work is also compared with the work of Minaee et al. as depicted in *table 5*.

Table 5: Performance comparison on FER13 dataset

Overall Accuracy (in %)	
Minaee et al. [36]	70.04
Our proposed method	94.08

5. CONCLUSION AND FUTURE WORKS

In this paper, a method is introduced to recognize emotion from different facial images with pose, occlusion, and illumination. From the past research, no such research has been done for the facial emotion recognition based on hybrid method. Despite of training is done in the dataset for still head poses and illuminations, our model is able to adapt all the variations like illumination, color, contrast, and head poses. That is, our hybrid model is able to give better results than traditional machine learning models. Our hybrid model is also able to produce good results with less training datasets in the publicly available datasets like EMOTIC, FER13, and FERG. Our model is able to detect emotion recognition with high accuracy and able to label each of them. The performance of our model for FER13 dataset is best as compare to FERG and EMOTIC datasets. In future, we will incorporate more deep learning methods to improve the results and also try to conduct some more experiments on other available datasets.

REFERENCES

[1] Jane, E.; Jackson, H.J.; Pattison, P.E. Emotion recognition via facial expression and affective prosody in schizophrenia: A methodological review. Clin. Psychol. Rev. 2002, 22, 789–832.

[2] Deepali, A.; Colburn, A.; Faigin, G.; Shapiro, L.; Mones, B. Modeling stylized character expressions via deep learning. In Asian Conference on Computer Vision; Springer: Cham, Switzerland, 2016; pp. 136–153.

[3] Chloé, C.; Vasilescu, I.; Devillers, L.; Richard, G.; Ehrette, T. Fear-type emotion recognition for future audio-based surveillance systems. Speech Commun. 2008, 50, 487–503.

[4] Rahul Mayur, Yadav Vikash et al, "Zernike Moments based Facial Expression Recognition using Two staged Hidden Markov Model", Advances in Computer Communication & Computational Sciences Proceedings of IC4S-2018", Vol. 924, pp. 661-670, May 22, 2019.

[5] N. Tiwari, S. Padhye, Analysis on the generalization of Proxy Signature, Security and Communication Network, Wiley, 2013 Vol. 6, pp. 549-556

[6] Meng, Q.; Hu, X.; Kang, J.; Wu, Y. On the effectiveness of facial expression recognition for evaluation of urban sound perception. Sci. Total Environ. 2020, 710, 135484.

[7] Marco, L.; Carcagnì, P.; Distante, C.; Spagnolo, P.; Mazzeo, P.L.; Rosato, A.C.; Petrocchi, S. Computational assessment of facial expression production in ASD children. Sensors 2018, 18, 3993.

[8] Ali, M.; Chan, D.; Mahoor, M.H. Going deeper in facial expression recognition using deep neural networks. In Proceedings of the IEEE 2016 IEEEWinter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, USA, 7–10 March 2016.

[9] Petrantonakis, C.P.; Hadjileontiadis, L.J. Emotion recognition from EEG using higher order crossings. IEEE Trans. Inf. Technol. Biomed. 2010, 14, 186–197.

[10] Wu, C.-H.; Chuang, Z.-J.; Lin, Y.-C. Emotion recognition from text using semantic labels and separable mixture models. ACM Trans. Asian Lang. Inf. Process. TALIP 2006, 5, 165–183.

[11] Courville, P.L.C.; Goodfellow, A.; Mirza, I.J.M.; Bengio, Y. FER-2013 Face Database; Universit de Montreal: Montréal, QC, Canada, 2013.

[12] LeCun, Y. Generalization and network design strategies. Connect. Perspect. 1989, 119, 143–155.

[13] Rahul Mayur, Yadav Vikash et al, "Gabor Filter and ICA based Facial Expression Recognition using Two Layered Hidden Markov Model", Advances in Computational Intelligence and Communication Technology Proceedings of CICT-2019", Vol. 1086, pp. 511-518, June 19, 2020.

[14] Singh Swarnima & Yadav Vikash, "Face Recognition using HOG Feature Extraction and SVM Classifier", International Journal of Emerging Trends in Engineering Research (IJETER), Vol. 8, No. 9, pp. 6437-6440, September 2020.

[15] Cohn, F.J.; Zlochower, A. A computerized analysis of facial expression: Feasibility of automated discrimination. Am. Psychol. Soc. 1995, 2, 6.

[16] Ekman, P.; Friesen, W.V. Constants across cultures in the face and emotion. J. Personal. Soc. Psychol. 1971, 17, 124.

[17] Friesen, E.; Ekman, P.; Friesen, W.; Hager, J. Facial Action Coding System: A Technique for the Measurement of Facial Movement; Psychologists Press: Hove, UK, 1978.

[18] Alex, K.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. 2012, 25, 1097–1105.

[19] Pooya, K.; Paine, T.; Huang, T. Do deep neural networks learn facial action units when doing expression recognition? In Proceedings of the IEEE International Conference on Computer Vision Workshops, Santiago, Chile, 7–13 December 2015.

[20] Liu, P.; Han, S.; Meng, Z.; Tong, Y. Facial expression recognition via a boosted deep belief network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1805–1812.

[21] Barsoum, E.; Zhang, C.; Ferrer, C.C.; Zhang, Z. Training deep networks for facial expression recognition with crowd-sourced label distribution. In Proceedings of the 18th ACM International Conference on Multimodal Interaction, Tokyo, Japan, 12–16 November 2016.

[22] Han, Z.; Meng, Z.; Khan, A.-S.; Tong, Y. Incremental boosting convolutional neural network for facial action unit recognition. Adv. Neural Inf. Process. Syst. 2016, 29, 109–117.

[23] Meng, Z.; Liu, P.; Cai, J.; Han, S.; Tong, Y. Identity-aware convolutional neural network for facial expression recognition. In Proceedings of the IEEE International Conference on Automatic Face & Gesture Recognition, Washington, DC, USA, 30 May–3 June 2017; pp. 558–565. Sensors 2021, 21, 3046 16 of 16

International Journal of
Electrical and Electronics Research (IJEER)
Research Article | Volume 10, Issue 1 | Pages 18-22 | e-ISSN: 2347-470X

[24] Marrero Fernandez, P.D.; Guerrero Pena, F.A.; Ren, T.; Cunha, A. Feratt: Facial expression recognition with attention net. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern RecognitionWorkshops, Long Beach, CA, USA, 16–20 June 2019.

[25] Wang, K.; Peng, X.; Yang, J.; Lu, S.; Qiao, Y. Suppressing uncertainties for large-scale facial expression recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020.

[26] Peng, K.W.; Yang, X.; Meng, D.; Qiao, Y. Region attention networks for pose and occlusion robust facial expression recognition. IEEE Trans. Image Process. 2020, 29, 4057–4069.

[27] Gan, Y.; Chen, J.; Yang, Z.; Xu, L. Multiple attention network for facial expression recognition. IEEE Access 2020, 8, 7383–7393.

[28] Arpita, G.; Arunachalam, S.; Balakrishnan, R. Deep self-attention network for facial emotion recognition. Proc. Comput. Sci. 2020, 17, 1527–1534.

[29] Shan, L.; Deng,W. Deep facial expression recognition: A survey. IEEE Trans. Affect. Comput. 2020.

[30] Donahue, J., Anne Hendricks, L., Guadarrama, S., Rohrbach, M., Venugopalan, S., Saenko, K., & Darrell, T. 2015. Longterm recurrent convolutional networks for visual recognition and description. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.2625-2634.

[31] R. Kosti, J.M. Álvarez, A. Recasens and A. Lapedriza, "Context based emotion recognition using emotic dataset", IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 2019.

[32] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee, Y. Zhou, C. Ramaiah, F. Feng, R. Li, X. Wang, D. Athanasakis, J. Shawe-Taylor, M. Milakov, J. Park, R. Ionescu, M. Popescu, C. Grozea, J. Bergstra, J. Xie, L. Romaszko, B. Xu, Z. Chuang, and Y. Bengio. Challenges in representation learning: A report on three machine learning contests. Neural Networks, 64:59--63, 2015. Special Issue on "Deep Learning of Representations"

[33] D. Aneja, A. Colburn, G. Faigin, L. Shapiro, and B. Mones. Modeling stylized character expressions via deep learning. In Proceedings of the 13th Asian Conference on Computer Vision. Springer, 2016.

[34] Grahlow M, Rupp CI, Derntl B (2022) The impact of face masks on emotion recognition performance and perception of threat. PLoS ONE 17(2): e0262840. https://doi.org/10.1371/journal.pone.0262840.

[35] Wang Kay Ngai, Haoran Xie, Di Zou, Kee-Lee Chou, Emotion recognition based on convolutional neural networks and heterogeneous bio-signal data sources, Information Fusion, Volume 77, 2022, Pages 107-117, ISSN 1566-2535, https://doi.org/10.1016/j.inffus.2021.07.007.

[36] Minaee, S.; Minaei, M.; Abdolrashidi, A. Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network. Sensors 2021, 21, 3046. https:// doi.org/10.3390/s21093046.

[37] Shashank M Gowda and H N Suresh (2022), Facial Expression Analysis and Estimation Based on Facial Salient Points and Action Unit (AUs). IJEER 10(1), 7-17. DOI: 10.37391/IJEER.100102.