ISSN: 2454-9940

# INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 18, Issue 4, 2024

# Efficient Emotion Recognition with a New Hybrid Approach Using Deep Learning

Dr. N.V. Subba Rao

Professor, Department of EEE

Anu Bose Institute of Technology for Women's, Paloncha, Telangana, India

Email: nvsubbaraonv@gmail.com

**ABSTRACT-** Due to huge intra-class fluctuations, facial expression detection has been a highly popular subject for academics in the last several decades, but it is also a very difficult and complicated endeavour. Current frameworks for this kind of challenge rely on methods like as Gabor filters, PCA, and ICA, which are followed by classification approaches that are trained using provided videos and pictures. While these frameworks do a good job with static photographs with consistent faces and lighting, they struggle when faced with dynamic images taken in a variety of settings. Several studies have presented deep learning frameworks for face emotion identification in recent years. Despite its efficacy, their research is consistently lacking. Our study presents a hybrid strategy that combines RNN and CNN. This approach successfully retrieves relevant elements from the provided database and achieves excellent results on datasets such as EMOTIC, FER-13, and FERG. Our hybrid methodology also achieves promising accuracy with these datasets, as we can demonstrate.

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

## 1. INTRODUCTION

People can't help but communicate with one another via their facial expressions. There are a lot of ways to put them to use that regular people would miss. This means that the provided tools can identify and recognise any indicator that comes after or before them. In recent decades, there has been a surge in interest in human facial emotion recognition due to the growing need for emotion identification in various domains such as medicine, animation, security, human-computer interaction, and the diagnosis of autism spectrum disorders in children and urban sound perception. Several features, such as facial expressions [8], EEG [9], and text [10], might be used to process the emotion detection in faces. Because they are observable and comprise a variety of characteristics, facial expressions are highly used in these features for effective emotion identification. On top of that, face collections are a breeze [11]. Using deep learning, recognition performance were much improved in the early years [12]. Facial emotion identification relies on a number of key essential aspects.

structure [13,14]. However, specific parts of the face, such as the eyes, nose, and lips, are all that are needed for facial emotion detection systems; other parts of the face, such as the hair and forehead, don't really contribute much to the identification of emotions [15]. We may conclude that the majority of our face emotion detection systems rely only on

one area and not on any of the others. We presented the hybrid technique for effective emotion identification in this study. The feature extraction strategy in this method uses CNN and RNN, while the classification technique uses SVM, both of which are deep learning techniques. Our investigation mostly uncovered:

(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,

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.

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 18, Issue 4, 2024

Prior research on face emotion identification has mostly concentrated on the conventional 2-step method that makes use of machine learning [34]. Step one involves extracting essential features from the picture using tools like Gabor filters, LBP, LMSP, and Zernike moments etch; step two involves classifying the image using tools like random forest, SVM, and KNN [35]. These methods work OK with relatively small datasets, but they struggle to keep up with increasingly large datasets. Sunglasses, half faces, occlusions, and dynamic backgrounds are some of the new picture difficulties and challenges.

Thanks to deep learning's impressive track record, particularly with convolutional neural networks (CNNs) for fast picture classification and other computer vision problems, a number of research groups are experimenting with deep learning concepts to identify face emotions [18]. Using the zero-bias CNN on the Toronto face dataset (TFD) and the expanded Cohn-kanade dataset (CK+), Khorrami et al. developed a CNN-based model for improved face emotion recognition accuracy [19]. A deep learning framework for animated character-based face expression identification was presented by Aneja et al. Human face network models were trained by them. They were also successful in training both animated and human faces to recognise and respond to certain expressions and gestures [2]. For face emotion identification, Mollahosseini et al. [8] suggested a neural network architecture with a single pooling layer, four inception layers, and two convolutional layers. Combining classification and feature extraction into a single looping web accessing two portions for feedback is what Liu et al. called a hybrid system. The author achieved the highest level of current accuracy by applying a boosted deep belief network (BDBN) to JAFFE and CK+ [20]. A deep learning approach for crowdsourcing noisy label collection in truth pictures was suggested by Barsoum et al. [21]. The author used several cost algorithms for DCNN and a total of ten taggers to rename all of the images in the dataset, producing the best possible outcome. In order to improve the accuracy rate for the datasets of spontaneous pictures, Han et al. [22] developed an incremental boosting convolutional neural network (CNN) known as IB-CNN. The outcomes were greatest at that time when this strategy was used. Minimising changes in identification and emotion-based information was the goal of Meng et al.'s identity-aware convolutional neural network (CNN) (IA-CNN) [23]. In their work on emotion detection using an attention-based model, Fernandez et al. [24] laid up a comprehensive web structure. To effectively deal with ambiguity and avoid unclear face expression pictures, Want et al. [25] presented a methodology based on a self-cure based network. Additionally, self-cure networks lowered uncertainty from both sources: (1) a method for each training run that calculates itself over a short batch update the labels of the provided sample in the lowest ranking class using regularisation of ranking (2), a relabelling approach. An effective method for face expression identification in real-world scenarios with occlusion modification was presented by Wang et al. In order to learn the unique characteristics of the face area and occlusion in FER, they successfully implemented a Region Attention Network (RAN). Current research on FER includes a literature review [28], a self-attention network [27] based on deep learning, and multiple attention networks [26] that use facial emotion recognition. The presented studies all outperformed the current state of the art in facial emotion identification, but they didn't use a specific method for detecting expressions. This study aims to address this limitation by developing a system that combines support vector machines (SVMs) for emotion classification with recurrent neural networks (RNNs) and convolutional neural networks (CNNs) for feature extraction.

## 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.

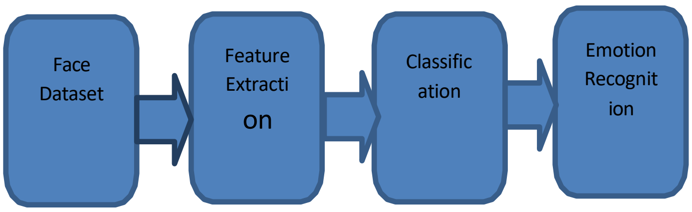


**Figure 1:** Facial Emotion Recognition System

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

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940
www.ijasem.org
Vol 18, Issue 4, 2024

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

Utilising the publicly available EMOTIC dataset, which includes 18,313 images annotated with 23,788 people, and FER-13, we can now generate some results. FER2013 includes approximately 30,000 facial images capturing distinct expressions, with a size of 48×48. The primary dataset can be further classified into seven types: zero=Angry, one=Disgust, two=Fear, three=Happy, four=Sad, five=Surprise, and six=Neutral. There are a minimum of 600 photos for the disgust facial emotion in the collection, compared to about 5,000 samples for each of the other classes. The FERG collection, for example, has 55,769 annotated photos of six cartoon characters' faces [32]. The seven cardinal emotions—surprise, melancholy, neutral, pleasure, wrath, disgust, and fear—are applied to every character [33]. In each scenario, we may train the model on a subset of the dataset, test it on the validation set, and determine its correctness using the 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 |

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

An approach to emotion recognition using position, occlusion, and lighting is presented in this study. There hasn't been any prior work on hybrid method-based face expression recognition. Even though the dataset is trained using static head postures and lighting, our model can handle all kinds of fluctuations in lighting, colour, contrast, and head positions. Put simply, our hybrid approach outperforms more conventional machine learning algorithms. Even with fewer training datasets, our hybrid model achieves respectable results on publically accessible datasets such as EMOTIC, FER13, and FERG. We have developed a model that can accurately identify and categorise different emotions. When compared to the FERG and EMOTIC datasets, our model performs the best on the FER13 dataset. We want to use other deep learning techniques in the future in an effort to enhance the findings and to expand our experimentation to other publicly accessible 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

[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.