International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Facial Emotion Recognition using Three-Layer ConvNet with Diversity in Data and Minimum Epochs

Rakshith M D[1], Harish H Kenchannavar[2], Umakant P Kulkarni[3]

Submitted: 10/09/2022 **Accepted:** 20/12/2022

Abstract: Human emotions can be identified by recognizing their facial expressions. This relates to the applications such as medical diagnostics, human emotional analysis, human-robot interaction, etc. This study presents a novel Convolution Neural Network (CNN) model for recognizing emotions from facial images. The proposed model, "ConvNet-3", recognizes the emotions such as disgust, anger, fear, happy, sad, surprise and neutral. The main focus of the proposed research is on training accuracy of the model in lesser number of epochs. The proposed model is trained on FER2013 dataset and its performance is evaluated. ConvNet-3 consists of 3 convolution layers and two fully connected layers. As illustrated in experimental results, the ConvNet-3 obtains training accuracy of 88% and validation accuracy of 61% on FER2013 which is better than existing models. In contrast it is observed that the presented model over fits on CK+48 dataset.

Keywords: *Convolutional Neural Network (CNN), Facial Expression, Emotions, Epochs, Accuracy.*

1. Introduction

The face is the important entity that provides minimal signals. These signals help in secured interaction between human and machine. Human intentions and intimate emotions can be inferred through facial expression recognition [1]. Human's true emotions can be determined through the facial expressions [2]. Static image based feature extraction methods were applied for performing computer vision tasks [3-5]. Dynamic based approaches consider temporal features [6-8]. Machine learning and deep learning approaches are drastically used for recognizing emotions through facial expressions [9]. Recent improvements in the deep learning have resulted in the CNN based models that provide promising results [10-12]. The generalized view of facial expression recognition system is depicted in Fig.1.

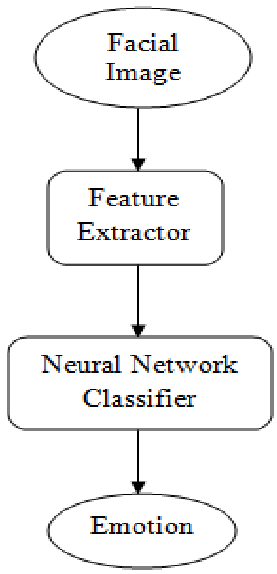

Fig. 1. Emotion recognition system

The proposed research focuses on recognizing the emotion from input facial image. The main aim of this study is to develop automated facial expression recognition (FER) system using Convolutional Neural Networks (CNNs). Traditional machine learning methods involves handcrafted features and is not reliable for the task of FER [13]. CNN-based models have obtained the best results for the recent FER tasks [11, 14]. The human faces can be categorized in real-time using the proposed model through the web camera. The contributions of the presented study to the

[1] *Department of Computer Science and Engineering, Canara Engineering College, Visvesvaraya Technological University, Karnataka, INDIA*
ORCID ID: 0000-0002-0196-531X
[2] *Gogte Institute of Technology, Belgaum, Karnataka, INDIA*
ORCID ID: 0000-0001-7369-0565
[3] *SDM College of Engineering & Technology, Dharwad, Karnataka, INDIA*
ORCID ID: 0000-0002-9452-4504
** Corresponding Author Email: rakshi.sdmit@gmail.com*

research community are as follows:

- The authors propose a compact CNN model for recognizing seven facial expressions. An effort has also been made to detect the facial expressions in real-time.
- Greater consistency is demonstrated by considering longer training sets.
- The proposed model provides a training accuracy of over 85% in a minimum number of epochs. This shows that model is well suited for the task of emotion recognition.
- The proposed model achieves accuracy better than the baseline 14% [15] in classifying seven distinct emotions.

2. Literature Survey

From many years the studies have been carried out on facial communication. But there is still room for carrying out experiments illustrating progress. Enhancing the precision of FER2013 [16] dataset is the key motive of researchers. The authors have applied CNN model to recognize seven critical emotions. The overall accuracy rate is 91%. But the identification rate is below 50% in classifying fear and disgust. In [17] the authors have recognized facial expressions using CNN. The proposed approach focused on mixed images chosen from multiple datasets. This enforces a major challenge in the domain of machine learning. A.christy et.al [18] proposed speech emotion recognition system that involves machine learning and deep learning techniques for performing the task of prediction and classification. In this study, CNN obtained a recognition accuracy of 78.20%. The authors carried out the task of FER on static images by employing CNN without explicitly performing feature extraction [19]. This approach when applied on FER2013 dataset provided a test accuracy of 61.7%. A novel hybrid approach [20] that uses EEG signals for emotion recognition provided an accuracy of 82.84%. Jung et.al [21] presented a hybrid deep neural networks that involves CNN and linked layers. This obtained an accuracy of 97.3% on CK+ dataset. Francesca Nonis et.al [22] proposed a study on problems in FER methods and 3D approaches to solve them. The average recognition accuracy varied from 60%-90%. An FER system presented in [23] shows that features extracted by using CNN are better than the handcrafted ones. A 3D video based emotion detection approach is discussed in [24]. This 3D based CNN model obtained an accuracy of 97.56% with cross validation on CK+ and 84.17% on Oulu-CASIA.

According to the authors the existing FER approaches exhibit certain drawbacks. Some of them are lack of model flexibility, capturing the feelings in complex situations, less number of participants, ineffective detection of EEG signals. These are the issues that need to be addressed further by enhancing user experience and functional applicability.

3. Proposed Work

In visual image processing and computer vision CNN is the most widely used deep neural network. Traditional algorithms used for FER face problems such as variance in location and light intensity. These problems can be solved by employing CNN algorithm for emotion recognition [25]. The proposed CNN model is structured sequentially as the sequential model has the capability to analyze and make predictions.

3.1 CNN Architecture

There are three layers of convolution and two layers which are fully connected layers in the presented model as shown in the Fig. 2. The functions of network are learnt in the convolution layer and also weights are updated. Fixed function is used by pooling layers to transform the activation. Through the loops the efficiency of convolution layer is inferred. Backpropagation algorithm is used to update the weights and minimize the loss function.

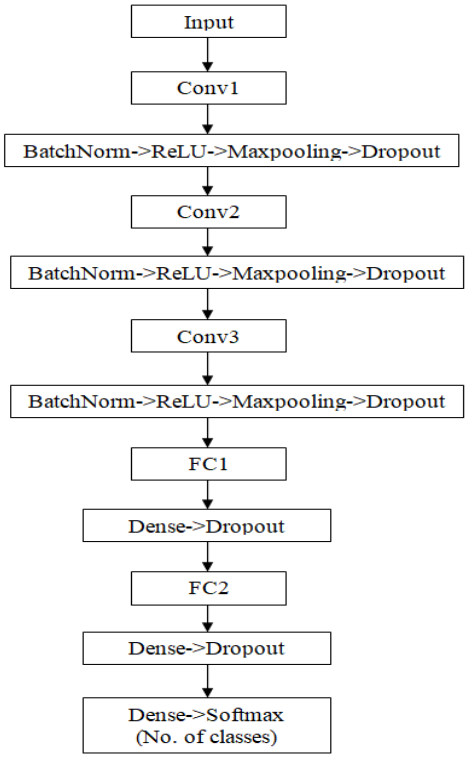

Fig. 2. Proposed CNN Stack

The process of fine-tuning in the pre-trained model replaces the fully connected layers with a new group of layers which are fully connected. By using backpropagation kernels in the convolution layers are also fine-tuned. Softmax activation function is used in the output layer for classification [26]. As shown in the Fig.2 the presented model involves ReLU layer, dropout layer and batch normalization. A dense layer is placed at the end of three convolution layers.

3.2 CNN Algorithm

This involves mathematical representation of the layers and their functions in the CNN.

Convolution layer

Input: Let $a^{[l-1]}$ be the image with size $(n_H^{[l-1]}, n_W^{[l-1]}, n_C^{[l-1]})$

S1: Perform padding p^l and stride s^l on the input image

S2: Initialize number of filters, $n_C^{[l]}$ where each $K^{(n)}$ has the dimension $(f^{[l]}, f^{[l]}, n_C^{[l-1]})$

S3: Define bias of the n^th convolution $b_n^{[l]}$

S4: Define activation function ψ^l

S5: Compute the size of output image based on stride value

$$n_{H/W}^{[l]} = \left\lfloor \frac{n_{H/W}^{l-1} + 2p^l - f^l}{s^{[l]}} + 1 \right\rfloor ; s > 0 \quad (1)$$

$$= n_{H/W}^{l-1} + 2p^l - f^l ; s = 0 \quad (2)$$

Output: Image $a^{[l]}$ with size $(n_H^{[l]}, n_W^{[l]}, n_C^{[l]})$

Pooling layer

Input: Let $a^{[l-1]}$ be the image with size $(n_H^{[l-1]}, n_W^{[l-1]}, n_C^{[l-1]})$

S6: Perform padding p^l and stride s^l on the input image

S7: Initialize the size of pooling filter $f^{[l]}$

S8: Perform pooling operation ϕ^l

S9: Compute the size of output image based on stride value

$$n_{H/W}^{[l]} = \left\lfloor \frac{n_{H/W}^{l-1} + 2p^l - f^l}{s^{[l]}} + 1 \right\rfloor ; s > 0$$
$$= n_{H/W}^{l-1} + 2p^l - f^l ; s = 0$$

Output: Image $a^{[l]}$ with size $(n_H^{[l]}, n_W^{[l]}, n_C^{[l]} = n_C^{[l-1]})$

Fully connected layer

Input: Vector $a^{[i-1]}$

S10: Compute the output at the j[th] node of i[th] layer

$$z_j^{[i]} = \sum_{l=1}^{n_{i-1}} w_{j,l}^{[i]} a_l^{[i-1]} + b_j^{[i]} \rightarrow a_j^{[i]} = \psi^{[i]}\left(z_j^{[i]}\right) \quad (3)$$

Output: Vector $a^{[i]}$

3.3 Dataset

The datasets used in the proposed study is FER2013 [27], CK+48 [32]. These are the most widely used databases for performing the task of FER. The training set in FER2013 consists of 28709 images and testing set involves 3589 images. In CK+48 the training set involves 735 images and testing set consists of 246 images. The distribution of expressions in the FER2013 dataset is depicted in the table 1.

Table 1. FER2013 Dataset Information

Sl.No	Emotion	Numeric code	No. of Images
1	Angry	0	4593
2	Disgust	1	547
3	Fear	2	5121
4	Happy	3	8989
5	Sad	4	6077
6	Surprise	5	4002
7	Neutral	6	6198

4. Results Analysis

The presented 3-layered CNN model was trained on the FER2013 database. Training accuracy of above 85% and validation accuracy of above 60% was observed. These accuracy results were obtained after 30 epochs. The comparison of proposed approach with the state-of-the art works is shown in table 2. This illustrates that the presented model is better than the other models for the task of FER.

Table 2. Comparative Analysis

Model	Accuracy (%)
VGG[28]	65-68
AlexNet[29]	55-88
Resnet [28]	72-74
CNN [31]	60-66
Proposed	75-88

When the proposed model is trained through each epoch it was observed that the accuracy of the model increased and the loss decreased. Training and validation accuracies obtained by the model on FER2013 are specified in the table 3. The training and validation accuracies obtained by the model on CK+48 are illustrated in table 4. It clearly exhibits overfitting.

Table 3. Accuracy per epoch on FER2013

	FER2013	
Epoch	Training Accuracy	Validation Accuracy
1	30.59	37.70
5	54.71	54.50
10	62.74	59.24
15	71.00	59.46
20	78.67	59.68
25	84.22	60.13
30	88.00	60.80

Table 4. Accuracy per epoch on CK+48

	CK+48	
Epoch	Training Accuracy	Validation Accuracy
1	24.76	13.82
5	60.95	19.92
10	80.95	15.04
15	90.34	17.07
20	94.83	10.16
25	94.97	6.10
30	96.05	6.50

Model fitting is determined by assessing training and validation accuracies. Model over-fitting occurs when there is large difference between the two. As shown in Fig.3, the training accuracy increases with increase in the epoch.

In Fig. 4, the training and validation losses are illustrated. The extent to which the model fits the training data is measured by training loss. Validation loss measures the extent to which the model fits the test data. The training and validation loss decreases as the epoch increases. When weights are updated, there is a decrease in the validation data.

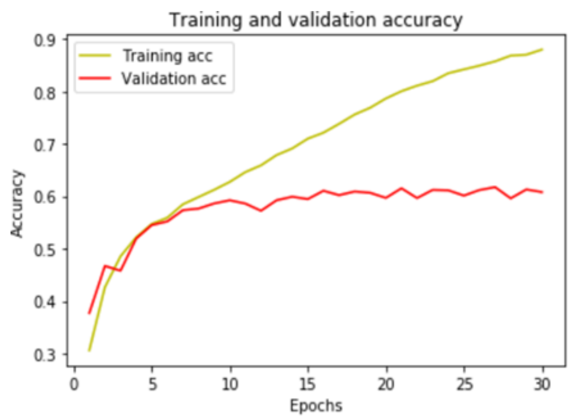

Fig. 3. Training and Validation accuracy per epoch

Fig. 4. Training and Validation loss per epoch

The confusion matrix obtained by the proposed model is depicted in Fig. 5. From the figure it is observed that among seven expressions the considerable accuracy is obtained for happy (84%) and surprise (72%) expressions. The diagonal matrix illustrates the number of images that are correctly classified.

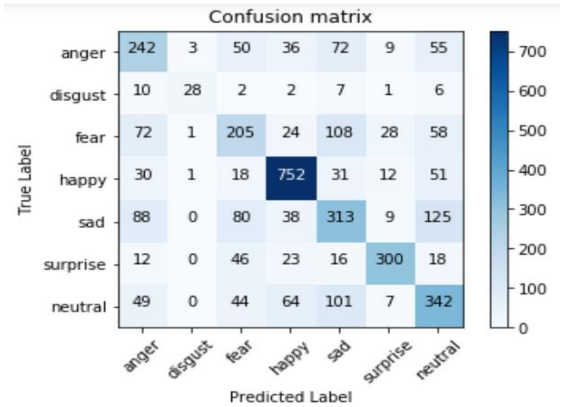

Fig. 5. Confusion matrix for the proposed model

5. Conclusion

In this article the authors have made an effort to recognize the following classes of emotions: angry, disgust, sad, surprise, happy, neutral and fear. As the proposed model fits the data perfectly in the case of FER2013 there is no overlapping between the classes. The presented model also provides a considerable accuracy of 88% with minimum number of epochs. But when the proposed model is applied on CK+48 dataset overfitting occurs. This is illustrated by high training and poor validation accuracies in recognizing the emotions. The proposed model can be evaluated in real-time to check its efficiency, so that it can be used in applications such as healthcare, entertainment, education, etc.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1]. L. Nwosu, H. Wang, J. Lu, I. Unwala, X. Yang, and T. Zhang, "Deep convolutional neural network for facial expression recognition using facial parts," in 2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress. IEEE, 2017, pp. 1318–1321.

[2]. R. Ekman, "What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)". Oxford University Press, USA, 1997.

[3]. N. Kumar and D. Bhargava, "A scheme of features fusion for facial expression analysis: A facial action recognition," *Journal of Statistics and Management Systems*, vol. 20, no. 4, pp. 693–701, 2017.

[4]. B. Yang, X. Xiang, D. Xu, X. Wang, and X. Yang, "3d palmprint recognition using shape index representation and fragile bits," *Multimedia Tools and Applications*, vol. 76, no. 14, pp. 357–375, 2017.

[5]. G. Tzimiropoulos and M. Pantic, "Fast algorithms for fitting active appearance models to unconstrained images," *International journal of computer vision*, vol. 122, no. 1, pp. 17–33, 2017.

[6]. Sudhakar, C. V., & Reddy, G. U. . (2022). Land use Land cover change Assessment at Cement Industrial area using Landsat data-hybrid classification in part of YSR Kadapa District, Andhra Pradesh, India. International Journal of Intelligent Systems and Applications in Engineering, 10(1), 75–86. https://doi.org/10.18201/ijisae.2022.270

[7]. H. Jung, S. Lee, J. Yim, S. Park, and J. Kim, "Joint fine-tuning in deep neural networks for facial expression recognition," in Proceedings of the IEEE international conference on computer vision, 2015, pp. 2983–2991.

[8]. G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 6, pp. 915–928, 2007.

[9]. X. Zhao, X. Liang, L. Liu, T. Li, Y. Han, N. Vasconcelos, and S. Yan, "Peak-piloted deep network for facial expression recognition," in European conference on computer vision. Springer, 2016, pp. 425–442.

[10]. H. Zhang, A. Jolfaei, and M. Alazab, "A face emotion recognition method using convolutional neural network and image edge computing," IEEE Access, vol. 7, pp. 159 081–159 089, 2019.

[11]. I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee et al., "Challenges in representation learning: A report on three machine learning contests," in International conference on neural information processing. Springer, 2013, pp. 117–124.

[12]. Z. Yu and C. Zhang, "Image based static facial expression recognition with multiple deep network learning," in Proceedings of the 2015 ACM on international conference on multimodal interaction, 2015, pp. 435–442.

[13]. S. E. Kahou, C. Pal, X. Bouthillier, P. Froumenty, Ç. Gülçehre, R. Memisevic, P. Vincent, A. Courville, Y. Bengio, R. C. Ferrari et al., "Combining modality specific deep neural networks for emotion recognition in video," in Proceedings of the 15th ACM on International conference on multimodal interaction, 2013, pp. 543–550.

[14]. C. Mayer, M. Eggers, and B. Radig, "Cross-database evaluation for facial expression recognition," *Pattern recognition and image analysis*, vol. 24, no. 1, pp. 124–132, 2014.

[15]. Y. Tang, "Deep learning using linear support vector machines," arXiv preprint arXiv:1306.0239, 2013.

[16]. C.-E. J. Li and L. Zhao, "Emotion recognition using convolutional neural networks," in Purdue Undergraduate Research Conference.63, 2019.

[17]. N. Christou and N. Kanojiya, "Human facial expression recognition with convolution neural networks," in Third International Congress on Information and Communication Technology. Springer, 2019, pp. 539–545.

[18]. Linda R. Musser. (2020). Older Engineering Books are Open Educational Resources. Journal of Online Engineering Education, 11(2), 08–10. Retrieved from http://onlineengineeringeducation.com/index.php/joee/article/view/41

[19]. S. M. González-Lozoya, J. de la Calleja, L. Pellegrin, H. J. Escalante, M. A. Medina, and A. Benitez-Ruiz, "Recognition of facial expressions based on cnn features," *Multimedia Tools and Applications*, pp. 1–21, 2020.

[20]. A. Christy, S. Vaithyasubramanian, A. Jesudoss, and M. A. Praveena, "Multimodal speech emotion recognition and classification using convolutional neural network techniques," *International Journal of Speech Technology*, vol. 23, pp. 381–388, 2020.

[21]. S. Singh and F. Nasoz, "Facial expression recognition with convolutional neural networks," in 2020 10th Annual Computing and Communication Workshop and Conference (CCWC). IEEE, 2020, pp. 0324–0328.

[22]. F. Wang, S. Wu, W. Zhang, Z. Xu, Y. Zhang, C. Wu, and S. Coleman, "Emotion recognition with convolutional neural network and eeg-based efdms," *Neuropsychologia*, vol. 146, p. 107506, 2020.

[23]. H. Jung, S. Lee, J. Yim, S. Park, and J. Kim, "Joint fine-tuning in deep neural networks for facial expression recognition," in Proceedings of the IEEE international conference on computer vision, 2015, pp. 2983–2991.

[24]. Gill, D. R. . (2022). A Study of Framework of Behavioural Driven Development: Methodologies, Advantages, and Challenges. International Journal on Future Revolution in Computer Science &Amp; Communication Engineering, 8(2), 09–12. https://doi.org/10.17762/ijfrcsce.v8i2.2068

[25]. F. Nonis, N. Dagnes, F. Marcolin, and E. Vezzetti, "3d approaches and challenges in facial expression recognition algorithms—a literature review," Applied Sciences, vol. 9, no. 18, p. 3904, 2019.

[26]. J. D. Bodapati and N. Veeranjaneyulu, "Facial emotion recognition using deep cnn based features," 2019.

[27]. J. Haddad, O. Lézoray, and P. Hamel, "3d-cnn for facial emotion recognition in videos," in International Symposium on Visual Computing. Springer, 2020, pp. 298–309.

[28]. A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in 2016 IEEE Winter conference on applications of computer vision (WACV). IEEE, 2016, pp. 1–10.

[29]. Anusha, D. J. ., R. . Anandan, and P. V. . Krishna. "Modified Context Aware Middleware Architecture for Precision Agriculture". International Journal on Recent and Innovation Trends in Computing and Communication, vol. 10, no. 7, July 2022, pp. 112-20, doi:10.17762/ijritcc.v10i7.5635.

[30]. Z. Yu and C. Zhang, "Image based static facial expression recognition with multiple deep network learning," in Proceedings of the 2015 ACM on international conference on multimodal interaction, 2015, pp. 435–442.

[31]. I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee et al., "Challengesin representation learning: A report on three machine learning contests," in International conference on neural information processing. Springer, 2013, pp. 117–124.

[32]. A. Kumar and G. Garg, "Sentiment analysis of multimodal twitter data," *Multimedia Tools and Applications*, vol. 78, no. 17, pp. 24 103–24 119, 2019.

[33]. Y.-L. Wu, H.-Y. Tsai, Y.-C. Huang, and B.-H. Chen, "Accurate emotion recognition for driving risk prevention in driver monitoring system," in 2018 IEEE 7th Global Conference on Consumer Electronics (GCCE). IEEE, 2018, pp. 796–797.

[34]. P. Giannopoulos, I. Perikos, and I. Hatzilygeroudis, "Deep learning approaches for facial emotion recognition: A case study on fer-2013," in Advances in hybridization of intelligent methods. Springer, 2018, pp. 1–16.

[35]. A.Agrawal and N.Mittal, Using CNN for facial expression recognition: a study of the effects of kernel size and number of filters on accuracy, in: *Visual Computer*, 36(2), pp.405-412, 2020.