International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 05 | May – 2025 DOI: 10.55041/ISJEM03747
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Facial Emotion Detection Using Deep Learning

M. Vinay Kumar
Department of ECE
Geethanjali College of Engineering & Technology
Hyderabad, India
vinaykumarrao66@gmail.com

K. Jahnavi
Department of ECE
Geethanjali College of Engineering & Technology
Hyderabad, India
kasam.jahnavi05@gmail.com

B. Anvesh
Department of ECE
College of Engineering & Technology
Hyderabad, India
bushianvesh@gmail.com

O.V.P.R. Siva Kumar
Professor, Department of ECE Geethanjali
Geethanjali College of Engineering College
Hyderabad, India
ogirala.sivakumar@gmail.com

Abstract— Human emotion detection from images is one of the most significant and challenging research tasks in social communication. Deep learning (DL)-based emotion detection provides better performance than traditional image processing methods. This paper presents the design of an artificial intelligence (AI) system capable of detecting emotions through facial expressions. It discusses the procedure of emotion detection, which mainly involves three steps: face detection, feature extraction, and emotion classification. This paper proposes a convolutional neural network (CNN)-based deep learning architecture for emotion detection from images. The performance of the proposed method is evaluated using two datasets: Facial Emotion Recognition Challenge (FER-2013) and Japanese Female Facial Expression (JAFFE). The accuracies achieved by the proposed model are 70.14% and 98.65% for the FER-2013 and JAFFE datasets, respectively.

Index Terms—Artificially intelligence (AI), Facial emotion recognition (FER), Convolutional neural networks (CNN), Rectified linear units (ReLu), Deep learning (DL).

I. Introduction

Emotions play a vital role in human communication, often expressed through subtle facial cues that transcend linguistic and cultural barriers. The ability to automatically recognize these emotions from facial expressions has become increasingly important in domains such as human-computer interaction, security, education, and mental health diagnostics. Over the years, researchers have strived to develop robust and intelligent systems capable of interpreting these expressions with high accuracy, drawing inspiration from psychological theories and leveraging computational advancements.

The study of facial expressions dates back to the 19th century, notably with Charles Darwin's foundational work on the universality of emotions [4]. Building upon this, Ekman and Friesen [5] identified six basic emotions—happiness, sadness, anger, surprise, fear, and disgust—that are universally recognized across cultures. Early computational efforts to analyze facial expressions relied on hand-crafted features and geometric models [3], [7], but these approaches were often limited by their inability to generalize across variations in lighting, occlusions, and individual facial structures.

The advent of deep learning has significantly transformed the landscape of facial emotion recognition (FER), enabling end- to-end learning from raw image data and eliminating the need for manual feature engineering. Convolutional Neural Networks (CNNs), in particular, have demonstrated remarkable performance in visual recognition tasks [10], [11], fueled by large-scale annotated datasets such as ImageNet [9] and FER- 2013 [16]. These models automatically learn hierarchical feature representations, making them well-suited for the complexities of facial expression analysis.

Several surveys have comprehensively reviewed the progress in this domain. Li and Deng [1] highlight the evolution of FER from shallow models to modern deep architectures, while Fasel and Luettin [7] provide insights into the challenges of real- world deployment, such as inter-subject variability and spontaneous expression recognition. Moreover, hybrid approaches incorporating temporal dynamics, such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) models, have shown promise in capturing the temporal evolution of facial expressions [14].

Recent research continues to refine FER systems by exploring advanced architectures, transfer learning, and multimodal fusion. The integration of pre-trained models like AlexNet [10] and frameworks such as TFLearn [12] has simplified experimentation and accelerated progress. In addition, publicly available datasets such as CK+ [15] and tools like OpenCV's Haar cascades [13] have been instrumental in training and evaluating emotion recognition models.

In this paper, we propose a deep learning-based approach for facial emotion recognition, leveraging CNN architectures to automatically extract and classify emotional states from static facial images. Our method is trained and tested on benchmark

datasets, aiming to contribute to the growing field of affective computing with a robust and scalable FER system.

II. Related Work

Facial Emotion Recognition (FER) has undergone significant evolution over the past decades, transitioning from traditional handcrafted feature-based approaches to deep learning-driven techniques. Early studies in the field were grounded in psychology and behavioral science, laying the theoretical foundation for computational models. Darwin's seminal work highlighted the biological basis and universality of facial expressions [4], later supported by Ekman and Friesen's identification of six universal emotions—happiness, sadness, anger, fear, surprise, and disgust—across cultures [5].

Initial computational efforts in FER focused on extracting geometric and appearance-based features, such as Action Units (AUs) in the Facial Action Coding System (FACS) [3]. These handcrafted features were then classified using machine learning techniques such as Support Vector Machines (SVMs) and Hidden Markov Models (HMMs). Although these methods provided valuable insights, their performance was often constrained by their dependence on precise facial landmark detection and sensitivity to environmental variations [7].

With the advent of deep learning, particularly Convolutional Neural Networks (CNNs), the field experienced a paradigm shift. CNNs automatically learn spatial hierarchies of features from raw image data, bypassing the need for manual feature engineering. Li and Deng [1] provided a comprehensive survey of deep learning methods for FER, highlighting the superior performance of CNNs over traditional approaches. Similarly, Correa et al. [2] demonstrated the effectiveness of deep convolutional architectures in capturing emotional expressions with high accuracy in constrained settings. The use of large-scale datasets such as ImageNet [9] and FER- 2013 [16] has played a pivotal role in enabling deep FER models. Krizhevsky et al. [10] popularized deep CNNs through their groundbreaking work on ImageNet classification, while Krizhevsky and Hinton [8] contributed to early feature learning from small-scale images. These developments laid the groundwork for applying transfer learning in emotion recognition tasks, significantly improving model generalization.

Incorporating temporal dynamics into FER systems has also been explored. Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, are capable of modeling sequential data and have been used to capture the evolution of facial expressions over time [14]. Tian et al. [3] and Lucey et al. [15] utilized datasets like CK+ to recognize both static and dynamic emotional cues, demonstrating improved performance through temporal modeling.

Furthermore, advances in open-source tools and libraries have accelerated the development of FER systems. TFLearn [12], built on TensorFlow, provides a simplified interface for deep learning experiments, while OpenCV's Haar cascade classifiers [13] enable efficient face detection as a preprocessing step. Studies such as those by Lv et al. [11] have integrated these tools to construct effective deep learning pipelines for emotion classification.

Despite the progress, challenges remain in achieving robust real-time FER in unconstrained environments. Variability in facial expressions due to age, ethnicity, lighting, and occlusions continues to impact recognition accuracy. Nonetheless, ongoing research continues to explore solutions through deeper networks, multimodal data fusion, and improved dataset diversity.

III. Proposed Model

A. *Emotion Detection Using Deep Learning*

In this paper we use the deep learning (DL) open library "Keras" provided by Google for facial emotion detection, by applying robust CNN to image recognition [12]. We used two different datasets and trained with our proposed network and evaluate its validation accuracy and loss accuracy. Images extracted from given dataset which have facial expressions for seven emotions, and we detected expressions by means of an emotion model created by a CNN using deep learning. We have changed a few steps in CNN as compared to previous method using a keras library given by Google and also modified CNN architectutre which give better accuracy . We implemented emotion detection using keras with the proposed network.

B. *CNN Architecture*

The networks are program on top of keras, operating on Python, using the keras learn library. This environment reduces the code's complexity, since only the neuron layers need to be formed, rather than any neuron. The software also provides real-time feedback on training progress and performance, and makes the model after training easy to save and reuse. In CNN architecture initially we have to extract input image of 48*48*1 from dataset FERC-2013. The network begins with an input layer of 48 by 48 which matches the input data size parallelly processed through two similar models that is functionality in deep learning, and then concatenated for better accuracy and getting features of images perfectly as shown in Fig.1 which is our proposed model, Model-A. There are two submodels for the extraction of CNN features which share this input and both have same kernel size. The outputs from these feature extraction sub- models are flattened into vectors and concatenated into one long vector matrix and transmitted to a fully connected layer for analysis before a final output layer allows for classification.

This models contains convolutional layer with 64 filters each with size of [3*3], followed by a local contrast normalization layer, maxpooling layer, followed by one more convolutional layer, max pooling, flatten respectively. After that we concatenate two similar models and linked to a softmax output layer which can classify seven emotions. We use dropout of 0.2 for reducing over-fitting. It has been applied to the fully connected layer and all layers contain units of rectified linear units (ReLu) activation function.

First we are passing our input image to convolutional layer which consists of 64 filters each of size 3 by 3, after that it passes through local contrast normalization can remove average from neighbourhood pixels leads to get quality of feature maps, followed by ReLu activation function. Maximum pooling is used to reduce spatial dimension reduction so processing speed will increase. We are using concatenation for getting features of images (eyes, eyebrows, lips, mouth etc) perfectly so that prediction accuracy improved as compared to previous model.

Furthermore, it is followed by fully connected layer and softmax for classifying seven emotions. A second layer of maxpooling is added to reduce the number of dimensionality. Here, we use batch normalization, dropout, ReLu activation function, categorical cross entropy loss, adam optimizer, softmax activation function in ouput layer for seven emotion classification.

In JAFEE dataset, input image size is adjusted to that 128*128*3. The network starts with an input layer of 128 by 128 which matches the input data size parallely processed through two similar models as shown in Fig.1. Furthermore, it is concatenated and pass through one more softmax layer for emotion classification and all procedure is same as above.

In Model-B, previously proposed by Correa et al. [2], the network starts with a 48 by 48 input layer, which matches the size of the input data. This layer is preceded by one convolutional layer, a local contrast normalization layer, and one layer of maxpooling, respectively. Two more convolutional layers and one fully connected layer, connected to a softmax output layer, complete the network. Dropout has been applied to the fully connected layer and all layers contain units of ReLu.

IV. Experiment Details

We develop a network based on the concepts from [12], [13] and [14] to assess the two models (Model-A and ModelB) mentioned above on their emotion detection capability. This section describes the data used for training and testing, explains the details of the used data sets and evaluates the results obtained using two different datasets with two models.

A. Datasets

Neural networks, and particularly deep networks, needs large amounts of training data. In addition, the choice of images used for the training is responsible for a large part of the eventual model's performance. It means the need for a data set that is both high quality and quantitative. Several datasets are available for research to recognize emotions, ranging from a few hundred high resolution photos to tens of thousands of smaller images. The two, we will be debating in this work, are the Japanese Female Face Expression (JAFFE) [15], Facial Expression Recognition Challenge (FERC-2013) [16] which contains seven emotions like anger, surprise, happy, sad, disgust, fear, neutral.

The datasets primarily vary in the amount, consistency, and cleanness of the images. For example, the FERC-2013 collection has about 32,000 low-resolution images. It can also be noted that the facial expressions in the JAFFE (i.e. further extended as CK+) are posed (i.e. clean), while the FERC-2013 set displays "in the wild" emotions. This makes it harder to interpret the images from the FERC 2013 set, but given the large size of the dataset, a model's robustness can be beneficial for the diversity.

B. Training Details

We train the network using GPU for 100 epochs to ensure that the precision converges to the optimum. The network will be trained on a larger set than the one previously described in an attempt to improve the model even more. Training will take place with 20,000 pictures from the FERC-2013 dataset instead of 9,000 pictures. The FERC-2013 database also uses newly designed verification (2000 images) and sample sets (1000 images). It shows number of emotions in the final testing and validation set after training and testing our model. The accuracy will be higher on all validation and test sets than in previous runs, emphasizing that emotion detection using deep convolutional neural networks can improve the performance of a network with more information.

C. Results using Proposed Model

In emotion detection we are using three steps, i.e., face detection, features extraction and emotion classification using deep learning with our proposed model which gives better result than previous model. In the proposed method, computation time reduces, validation accuracy increases and loss also decreases, and further performance evaluation achieved which compares our model with previous existing model. We tested our neural network architectures on FERC-2013 and JAFFE database which contains seven primary emotions like sad, fear, happiness, angry, neutral, surprised, disgust.

Fig.2 shows the proportions of detected emotions in a single image of FER dataset. Fig.2(a) shows the image, whereas the detected emotion proportions are shown in Fig.2(b). It is clearly observable that neutral has higher proportion than other emotions. That means, the emotion detected for this image (in Fig.2(a)) is neutral. Similarly, Fig.3 show another image and corresponding emotion proportions. From Fig.3(b), it is observable that happy emotion has higher proportion than others. That suggests that image of Fig.3(a) detects happy emotions.

Similarly, performance is evaluated for all the test images of is evaluated for all the test samples of JAFFE dataset. When we are using JAFFE dataset we are getting validation accuracy of 98.65 percentage which is better than previous result and it takes less computational time per step.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 05 | May – 2025 DOI: 10.55041/ISJEM03747
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

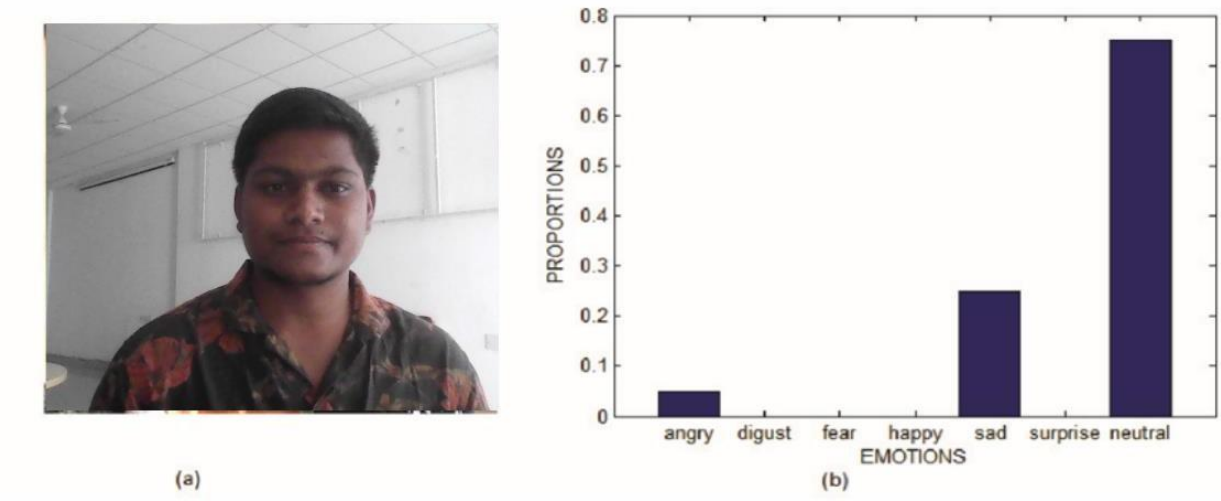


Fig. 2: (a) Image, (b) Proportion of emotions.

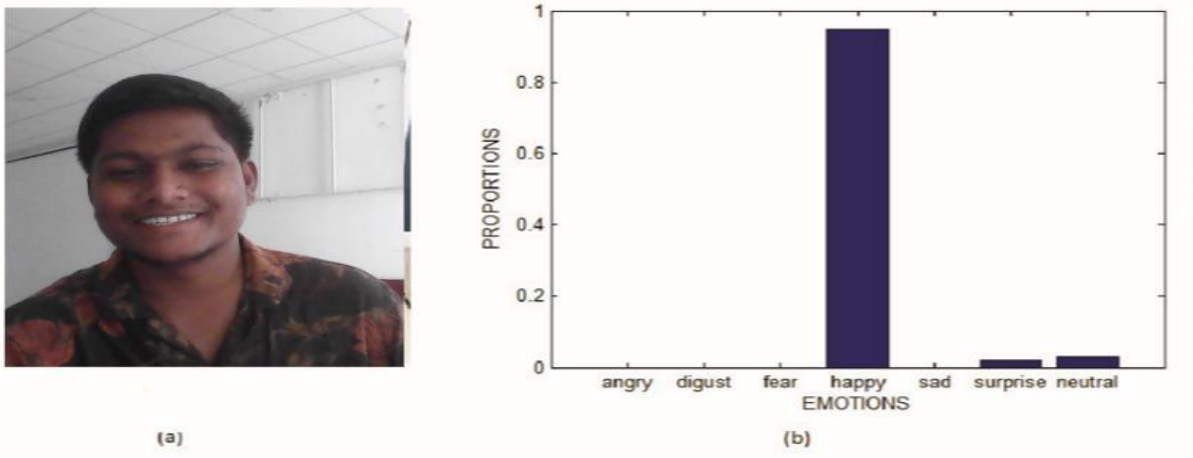
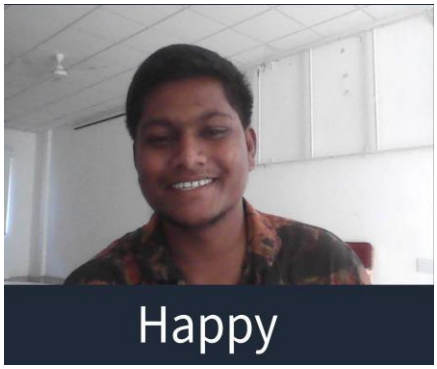


Fig. 3: (a) Image, (b) Proportion of emotions.

the dataset. We have achieved 95 percentage for happy, 75 percentage for neutral, 69 percentage for sad, 68 percentage for surprise, 63 percentage for disgust, 65 percentage for fear and 56 percentage for angry. On an average we are getting average accuracy of 70.14 percentage using our proposed model.
The confusion matrix of classification accuracy is shown in TABLE I. We get an average validation accuracy of 70.14 percentage using our proposed model in facial emotion detection using FER dataset.

Fig.4 shows the result of test sample related to surprise emotion from JAFFE dataset, and our proposed model also predicted the same emotion with reduced computation time as compared to previous existing model B. Similarly, performance



V. PERFORMANCE EVALUATION

In FER dataset we train on 32,298 samples which is validate on 3589 samples, and in JAFFE dataset we train 833 samples, which is validate on 148 samples for calculation of validation accuracy, validation loss, computational time per step upto to 100 and 50 epochs respectively shown in . The aim of the training step is to determine the correct configuration parameters for the neural network which are: number of nodes in the hidden layer (HL), rate of learning (LR), momentum (Mom), and epoch (Ep). Different combinations of these parameters have been tested to find out how to achieve the better recognition rate.

From Table II, it is observed that our proposed model shows 70.14% average accuracy compared to the 67.02% average accuracy reported in model B FOR FER dataset. In this case of JAFFE database, we achieved average accuracy 98.65% which is also higher than model B.

TABLE I: Confusion Matrix (%) for emotion detection using proposed model

Emotions	Angry	Sad	Happy	Disgust	Fear	Neutral	Surprise
Angry	56	12	3	9	8	11	1
Sad	10	69	2	6	9	2	2
Happy	0	0	95	0	0	3	2
Disgust	7	13	0	63	8	5	4
Fear	9	8	3	2	65	10	3
Neutral	2	1	8	1	7	75	6
Surprise	7	3	11	0	3	8	68
Average accuracy = 70.14 (%)							

VI. CONCLUSION

In this paper, we have proposed a deep learning based facial emotion detection method from image. We discuss our proposed model using two different datasets, JAFFE and FERC-2013. The performance evaluation of the proposed facial emotion detection model is carried out in terms of validation accuracy, computational complexity, detection rate, learning rate, validation loss, computational time per step. We analyzed our proposed model using trained and test sample images, and evaluate their performance compare to previous existing model. Results of the experiment show that the model proposed is better in terms of the results of emotion detection to previous models reported in the literature. The experiments show that the proposed model is producing state-of-the-art effects on both two datasets.

ACKNOWLEDGMENT

We would like to express our sincere gratitude to **Prof. O.V.P.R. Siva Kumar** for his valuable guidance, continuous support, and encouragement throughout the course of this research. His insightful suggestions and expertise in the field have been instrumental in shaping the direction and quality of our work on facial emotion recognition using deep learning. We are truly grateful for his mentorship and for providing us with the opportunity and resources to carry out this study successfully

REFERENCES

[1] S. Li and W. Deng, "Deep facial expression recognition: A survey," arXiv preprint arXiv:1804.08348, 2018.
[2] E. Correa, A. Jonker,M.Ozo, andR.Stolk, "Emotion recognition using deep convolutional neural networks," Tech. Report IN4015, 2016.
[3] Y. I. Tian, T. Kanade, and J.F.Cohn, "Recognizing action units for facial expression analysis," IEEE Transactions on pattern analysis and machine intelligence, vol. 23, no. 2, pp. 97–115, 2001.
[4] C. R. Darwin. The expression of the emotions in man and animals. John Murray, London, 1872.
[5] P. Ekman and W. V. Friesen. Constants across cultures in the face and emotion. Journal of personality and social psychology, 17(2):124, 1971.

[6] J. Nicholson, K. Takahashi, and R. Nakatsu. Emotion recognition in speech using neural networks. Neural computing applications, 9(4): 290–296, 2000.

[7] B. Fasel and J. Luettin. Automatic facial expression analysis: a survey. Pattern recognition, 36(1):259–275, 2003.

[8] A. Krizhevsky and G. Hinton. Learning multiple layers of features from tiny images, 2009.

[9] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pages 248–255. IEEE, 2009.

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pages 1097–1105, 2012.

[11] Y. Lv, Z. Feng, and C. Xu. Facial expression recognition via deep learning. In Smart Computing (SMARTCOMP), 2014 International Conference on, pages 303–308. IEEE, 2014.

[12] TFlearn. Tflearn: Deep learning library featuring a higher-level api for tensorflow. URL http://tflearn.org/.

[13] Open Source Computer Vision Face detection using haar cascades. URL http://docs.opencv.org/master/d7/d8b/tutorialpyfacedetection.html.

[14] P. J. Werbos et al., "Backpropagation through time: what it does and how to do it," Proceedings of the IEEE, vol. 78, no. 10, pp. 1550–1560, 1990.

[15] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on, pages 94–101. IEEE, 2010.

[16] Kaggle. Challenges in representation learning: Facial expression recognition challenge, 2013.