Journal of
Image Processing & Pattern Recognition Progress

ISSN: 2394-1995
Volume 12, Issue 2, 2025
May–August
DOI (Journal): 10.37591/JoIPPRP

https://journals.stmjournals.com/joipprp

Research JoIPPRP

Comparison and Analysis of Facial Emotion Detection Using Various Deep Learning Neural Networks

V. Anusuya[1,*], S. Sakkaravarthi[1], M. Rethinakumari[1], Sridhar S.[2]

Abstract

Facial emotion recognition employs Convolutional Neural Networks (CNNs), Residual Networks (ResNet), Long Short-Term Memory (LSTM) networks, and Deep Neural Networks (DNNs) to automatically identify various emotions, including disgust, anger, fear, happiness, sadness, surprise, and neutrality. This study utilizes transfer learning along with data preprocessing techniques such as rotation, flipping, brightness adjustment, and enhancement methods. Traditional machine learning models achieve an accuracy range of 45 to 50%. In contrast, our proposed CNN and DNN models show improved accuracy, reaching 65 and 62%, respectively. Additionally, we introduce a hybrid model combining ResNet and LSTM architecture, which achieves an accuracy of 72%. Following this, we conduct a comparative analysis of the accuracy and loss for each model. Our findings indicate that while CNNs initially demonstrated higher accuracy than the hybrid ResNet-LSTM model, the hybrid model ultimately surpassed all others in total classification accuracy. The performance metrics used in this analysis include recall and precision.

Keywords: Facial emotion recognition (FER), neural networks, convolutional neural networks (CNNs), residual networks (ResNet), long short-term memory (LSTM)

INTRODUCTION

Facial emotion recognition is crucial in the fields of computer vision and artificial intelligence. It involves identifying and interpreting human emotions through facial expressions. Convolutional Neural Networks (CNNs) and Residual Networks (ResNets) are highly effective for tasks that require learning spatial hierarchies of features. Meanwhile, Long Short-Term Memory (LSTM) networks and Deep Neural Networks (DNNs) are utilized for learning from temporal sequences and extracting deep features. The Work process of facial emotion recognition is mentioned in Figure 1.

***Author for Correspondence**
V. Anusuya
E-mail: anusuyav@ritrjpm.ac.in

[1]Associate Professor, Department of Information Technology, Ramco institute of Technology, Rajapalayam, Tamil Nadu, India
[2]Student, Third year, Department of Information Technology, Ramco institute of Technology, Rajapalayam, Tamil Nadu, India

Received Date: March 06, 2025
Accepted Date: April 06, 2025
Published Date: April 28, 2025

Citation: V. Anusuya, S. Sakkaravarthi, M. Rethinakumari, Sridhar S. Comparison and Analysis of Facial Emotion Detection Using Various Deep Learning Neural Networks. Journal of Image Processing & Pattern Recognition Progress. 2025; 12(2): 37–42p.

The FER2013 dataset, commonly utilized for evaluating emotion recognition models, consists of seven emotion categories: Anger, Disgust, Fear, Happiness, Sadness, Surprise, and Neutral, all captured under various conditions.

RELATED WORK

A system for automatic emotion recognition has been created using facial analysis techniques that integrate Local Binary Pattern (LBP) and Histogram of Oriented Gradients (HOG) features. This system utilizes two methods for dimensionality reduction: Principal Component Analysis (PCA) and Locally Linear Embedding (LLE). For the classification task, a multi-class Support Vector

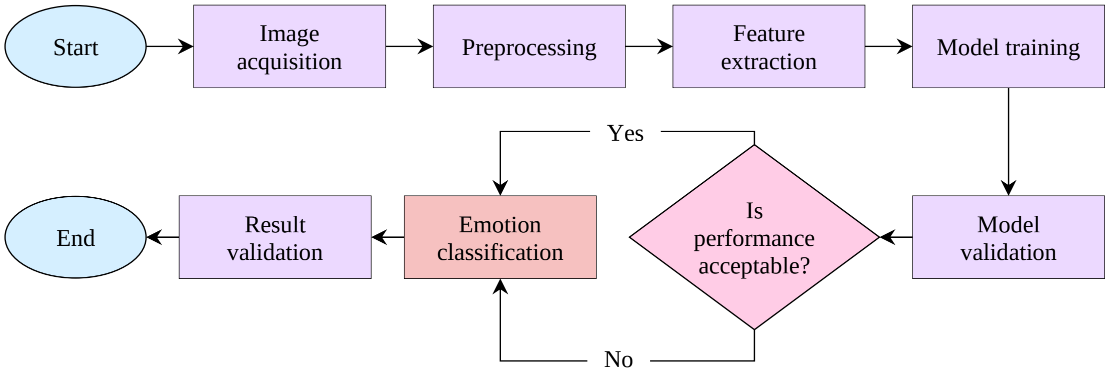

Figure 1. Work process of facial emotion recognition.

Machine (SVM) is selected because of its excellent ability to generalize and differentiate between six fundamental emotions. The effectiveness of the system is assessed using three established facial expression datasets: Japanese Female Facial Expression (JAFFE), Karolinska Directed Emotional Faces (KDEF), and Radboud Faces Database (RaFD), resulting in impressive recognition rates of 97.16, 90.12, and 95.54%, respectively [1].

A deep learning-based method is introduced to assess the real-time engagement of online learners by examining their facial emotions. The system categorizes students' facial expressions to identify their emotions during online learning sessions, which are then utilized to calculate an Engagement Index (EI) that predicts two states of engagement: “Engaged” and “Disengaged”. Several deep learning models, such as Inception-V3, VGG19, and ResNet-50, are tested and compared to find the most effective model for detecting engagement in real-time. The proposed system achieves accuracies of 89.11, 90.14, and 92.32% for Inception-V3, VGG19, and ResNet-50, respectively, using both benchmark datasets and a custom dataset. ResNet-50 shows the best performance, reaching 92.32% accuracy for classifying facial emotions in real-time learning settings [2]. Automatic systems that recognize emotions and facial expressions are being used in a range of areas, including healthcare, security, advertising, and marketing, among others [3].

Facial Emotion Recognition (FER) techniques that utilize deep learning and artificial intelligence (AI) have been improved with edge modules to enhance efficiency and enable real-time processing. These techniques are tailored for edge devices like smartphones and Raspberry Pi, with detailed discussions on various performance evaluation metrics for FER systems [4].

The proposed system contributed to the field by employing Convolutional Neural Networks (CNNs) to identify emotions and sentiment polarities in social media data. Their research showcased the ability of CNNs to categorize sentiment into positive, negative, and neutral based on facial expressions found in user-uploaded images, offering valuable insights into public opinion on social media platforms [5].

The study also emphasized the effectiveness of CNN models in handling large and diverse datasets. Additionally, the research examined the challenges that mask-wearing during the COVID-19 pandemic created for Facial Emotion Recognition (FER). It highlighted the necessity of adapting FER models to these new conditions, as masks covered essential facial features, which significantly impacted model performance.

The authors suggested modifications to CNN architectures, concentrating on the use of upper-face cues and enhancing pre-processing techniques to maintain high recognition accuracy despite the challenges posed by mask-wearing [6].

An Anti-Aliased Deep Convolutional Network (AA-DCN) model has been created to explore how anti-aliasing can improve the accuracy of facial emotion recognition. This model is capable of identifying eight different emotions from image data, leading to a significant enhancement in emotion recognition while minimizing aliasing artifacts that can occur due to down-sampling layers [7].

A new framework for recognizing facial expressions has also been introduced, which employs a hybrid model that merges Convolutional Neural Networks (CNNs) with a Support Vector Machine (SVM) classifier to analyze dynamic facial expression data. This system successfully classifies seven facial expressions: anger, contempt, disgust, fear, happiness, sadness, and surprise, using the CK+ database, and six expressions: anger, disgust, fear, happiness, sadness, and surprise, on the BU4D database. The proposed system achieves an impressive recognition accuracy of 99.69% on the CK+ database and 94.69% on the BU4D database [8].

This study presents a method for extracting features in facial emotion recognition by leveraging the deep residual network ResNet-50. It integrates the convolutional neural network (CNN) approach with ResNet-50, which has shown remarkable performance in multi-classification tasks throughout the facial expression recognition learning process [9].

Using the Haar Cascade classifier to crop faces and concentrate on the region of interest, we propose that this method will facilitate faster convergence by eliminating the need to analyze the entire image for facial expressions. Furthermore, we implement label smoothing and assess its effects on the CK+, KDEF, and RAF databases. The ResNet model exemplifies a neural network architecture, with label smoothing demonstrating an enhancement in recognition accuracy of up to 0.5% on the CK+ and KDEF databases [10].

PROPOSED METHODOLOGY

Real-Time Emotion Recognition

After capturing a video frame using a webcam or camera, the following steps are applied:

Preprocessing

Convert the frame to grayscale if needed, resize it to the required input dimensions (for example, 48×48 pixels), and normalize the pixel values.

- *Feature extraction:* Pass the pre-processed frame through the feature extraction layers of the trained models, such as a CNN, Resnet, LSTM and DNN.
- *Prediction:* Pass the extracted features into the model's fully connected layers to predict the emotion class. Show the predicted emotion on the user interface in real-time, offering immediate feedback on the emotions identified from the video frame.

Data Preprocessing

Convert the frame to grayscale, if necessary, resize it to the required input size (for example, 48×48 pixels), and normalize the pixel values. Normalize the pixel values from a range of 0 to 255 down to 0 to 1. This adjustment aids the neural network in converging more efficiently and stabilizes training by making gradient updates more predictable.

Data Augmentation

Introduce transformations to the data such as:

- *Rotation:* Slightly rotate images at random (e.g., –10 to +10°).
- *Flipping:* Flip images horizontally to add more variability.
- *Brightness adjustment:* Randomly change the brightness.

MODEL EVALUATION

In the case of models, the metrics such as the confusion matrix, F1-Score are fine. A confusion matrix is an N×M matrix of the actual classes against predicted classes for which the actual values are known, aiming to show how well a classification model works. F1 Score is a statistical measure defined as the harmonic mean of precision and recall as it provides useful information for recognition systems with unbalanced classes.

RESULTS AND DISCUSSION

In this segment, an in-depth analysis is made on the performance metrics of four deep learning models: Convolutional Neural Networks (CNNs), Residual Networks (ResNets) Long Short-Term Memory (LSTM) networks, and Deep Neural Networks (DNNs), in the task of facial emotion recognition on the FER2013 dataset. The evaluation is presented through loss as well as accuracy metrics in Figure 2, which are shown against the respective accuracy and loss comparison plots.

Accuracy Analysis

The models' outcomes are discussed. By the end of the training, the Convolution Neural Network model achieves a 65% validation accuracy. ResNet networks outperform standard CNNs and gain a validation accuracy score of 60% which makes CNNs more favorable for complex datasets such as FER2013.

LSTMs are not built for spatial data such as images, they work with sequential data. Their lower accuracy in this is caused by their greater focus on temporal dependencies which is not the case with still images. It attains 50% as validation accuracy. Deep Neural Networks (DNNs) are composed of fully connected layers, so, DNNs are not preferred for image data. Their lack of convolutional layers makes them incapable of capturing spatial features which results in their inability to perform facial emotion recognition, with validation accuracy around 40%.

Loss Analysis

The training loss for both CNN and ResNet is close to 1.0, suggesting that both models are learning well, although ResNet has a slight advantage because of its deeper architecture. In contrast, LSTMs and DNNs show higher training loss, with DNNs performing the worst. ResNet's performance is marginally better than that of CNN, which underscores its effectiveness in practical applications. On the other hand, LSTMs and DNNs exhibit considerable fluctuations in validation loss, indicating their instability and limited ability to generalize. Table 1 represents the Validation Accuracy and F1-Score.

F1 Score: F1-Score =2×(Precision × Recall)/(Precision + Recall)

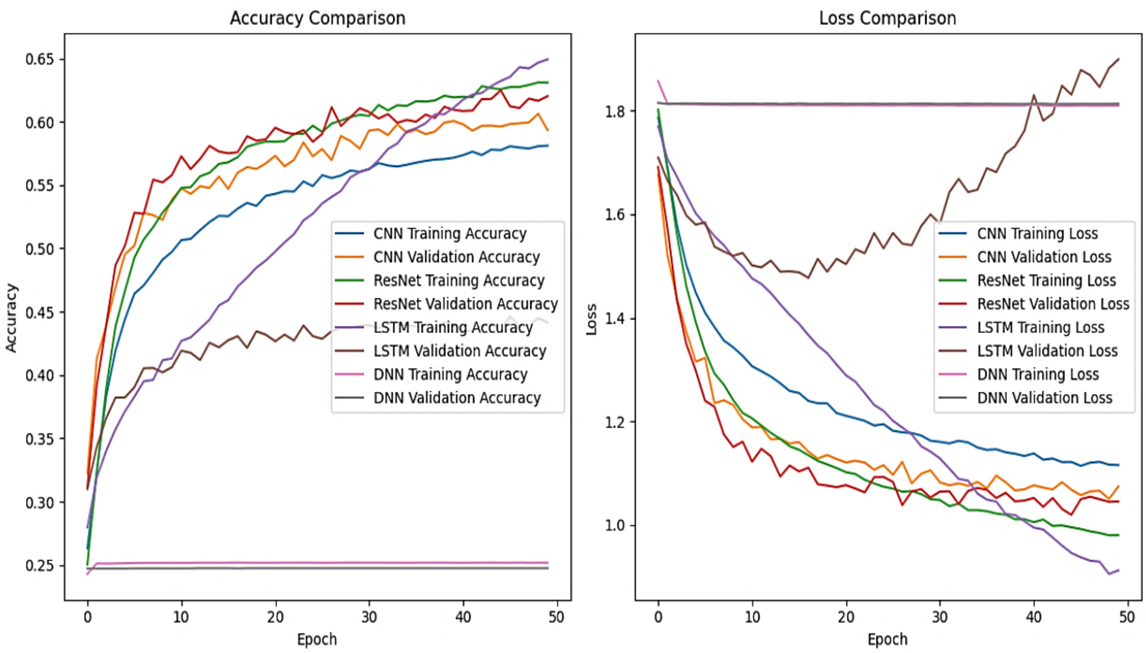

Figure 2. Accuracy and loss comparison.

Table 1. Validation accuracy and F1-Score.

Model	Validation accuracy	F1-Score
CNN	65%	66%
ResNet	60%	61.50%
LSTM	50%	52.50%
DNN	40%	42.50%

Comparison Analysis

In this evaluation, we examined the performance of various deep learning models: CNN, ResNet, LSTM, and DNN, on the task of recognizing facial emotions using the FER2013 dataset. Our findings indicate that CNNs and ResNets outperformed LSTM and DNN models, particularly regarding validation accuracy and generalization capabilities. CNNs achieved the highest validation accuracy at around 65%, closely followed by ResNets, which reached about 60% accuracy thanks to their residual learning architecture. In contrast, LSTMs and DNNs faced challenges with image-based tasks, recording validation accuracies of 50 and 40%, respectively, and showing higher loss values.

These results highlight the effectiveness of CNNs and ResNets in facial emotion recognition, as they are particularly adept at learning spatial features from images. Conversely, LSTM’s focus on sequential data processing and DNN’s fully connected layers were less effective for image data, emphasizing the need to choose architectures specifically designed for spatial data in image classification tasks.

For future research, further optimization of these models could be pursued, including hybrid architectures or the application of transfer learning to boost performance. Additionally, it will be crucial to address model robustness and fairness in emotion recognition systems to ensure reliable deployment in real-world scenarios.

CONCLUSION

In this evaluation, we assessed the performance of several deep learning models: CNN, ResNet, LSTM, and DNN, on the task of facial emotion recognition using the FER2013 dataset. Our results show that CNNs and ResNets outperformed LSTM and DNN models, especially in terms of validation accuracy and generalization capabilities. CNNs achieved the highest validation accuracy of approximately 65%, closely followed by ResNets, which reached around 60% accuracy due to their residual learning architecture. In contrast, LSTMs and DNNs struggled with image-based tasks, with validation accuracies of 50 and 40%, respectively, and exhibited higher loss values.

These findings underscore the effectiveness of CNNs and ResNets in facial emotion recognition, as they are well-suited for learning spatial features from images. On the other hand, LSTM’s sequential data processing and DNN’s fully connected layers proved less effective for image data, highlighting the importance of selecting architectures that are specifically designed for spatial data in image classification tasks.

For future work, optimization of these models can be explored further, including hybrid architectures or the use of transfer learning to enhance performance. It will also be essential to address model robustness and fairness in emotion recognition systems to ensure reliable real-world deployment.

REFERENCES

1. Yaddaden Y. An efficient facial expression recognition system with appearance-based fused descriptors. Intell Syst Appl. 2023 Feb 1; 17: 200166.
2. Gupta S, Kumar P, Tekchandani RK. Facial emotion recognition based real-time learner engagement detection system in online learning context using deep learning models. Multimed Tools Appl. 2023 Mar; 82(8): 11365–94.

3. Dias W, Andalo F, Padilha R, Bertocco G, Almeida W, Costa P, Rocha A. Cross-dataset emotion recognition from facial expressions through convolutional neural networks. J Vis Commun Image Represent. 2022 Jan 1; 82: 103395.
4. Sajjad M, Ullah FU, Ullah M, Christodoulou G, Cheikh FA, Hijji M, Muhammad K, Rodrigues JJ. A comprehensive survey on deep facial expression recognition: challenges, applications, and future guidelines. Alex Eng J. 2023 Apr 1; 68: 817–40.
5. Meena G, Mohbey KK, Indian A. Categorizing sentiment polarities in social networks data using convolutional neural network. SN Comput Sci. 2022 Mar; 3(2): 116.
6. Castellano G, De Carolis B, Macchiarulo N. Automatic facial emotion recognition at the COVID-19 pandemic time. Multimed Tools Appl. 2023 Apr; 82(9): 12751–69.
7. Elsheikh RA, Mohamed MA, Abou-Taleb AM, Ata MM. Improved facial emotion recognition model based on a novel deep convolutional structure. Sci Rep. 2024 Nov 23; 14(1): 29050.
8. Kim JC, Kim MH, Suh HE, Naseem MT, Lee CS. Hybrid approach for facial expression recognition using convolutional neural networks and SVM. Appl Sci. 2022 May 28; 12(11): 5493.
9. Li B, Lima D. Facial expression recognition via ResNet-50. International Journal of Cognitive Computing in Engineering (IJCCE). 2021 Jun 1; 2: 57–64.
10. Shehu HA, Sharif MH, Uyaver S. Facial expression recognition using deep learning. In AIP Conf Proc. 2021 Mar 2; 2334(1): 070003. AIP Publishing.

