OPEN ACCESS JOURNALS
Indiana Journal of Multidisciplinary Research
Abbriviate Tittle- Ind J Multi Res
ISSN *(Online)*- 2583-3820
Journal Homepage Link- https://indianapublications.com/Journals/IJMR
DOI: https://doi.org/10.5281/zenodo.12677782

INDIANA PUBLICATIONS
PRODUCTIVE AND QUALITY RESEARCH

Research Article

Volume-04|Issue-03|2024

Emotion Detection using Facial Expression – A comprehensive review

Haritha S[1], Sakshi Gupta[*2], Sara Mehraj[3], Deepak D J[4]

[1,2,3]Student, Department of Information Science, RV Institute of Technology and Management, Bangalore, Karnataka, India
[4]Assistant Professor, Department of Information Science, RV Institute of Technology and Management, Bangalore, Karnataka, India

Article History
Received: 20.05.2024
Accepted: 05.06.2024
Published: 30.06.2024

Citation
Haritha, S., Gupta, S., Mehraj, S., & Deepak, D. J. (2024). Emotion Detection using Facial Expression – A comprehensive review. *Indiana Journal of Multidisciplinary Research*, *4*(3), 174-178.

Abstract: Recent advancements in computer vision and artificial intelligence have significantly accelerated research in human emotion recognition through facial expressions. Central to this field of study are deep learning techniques that improve the accuracy of automated systems. This paper presents a comprehensive review of emotion recognition methods, examining convolutional neural networks as a potent solution for interpreting fundamental human emotions. We collate findings across various studies, highlighting the efficacy and challenges faced in emotion detection using facial features. Notably, research by (4) demonstrates the practical application of CNN using the publicly available FER2013 image dataset, emphasizing the importance of systematic approaches to facial emotion recognition. Moreover, as elucidated by Arora *et al.*, deep learning algorithms carry substantial promise for identifying complex emotional states, overcoming traditional computational limitations. Our study further investigates pre-processing techniques to enhance CNN performance, such as noise reduction and feature extraction, reinforcing the findings by (1) of their effect on classification accuracy for the seven emotions in the facial action coding system. Through systematic analysis and synthesis of the literature, we identify trends, methodologies, and applications within this field. Our research underscores the multidisciplinary impact of facial emotion recognition, spanning psychological assessment, human-computer interaction, and even medical diagnostics. The paper endeavours to contribute to the refinement of facial expression-based emotion recognition systems, aiming for a nuanced understanding of human affective states that surpasses existing computational paradigms.

Keywords: Emotion Recognition, CNN, Deep Learning, Facial Expression

INTRODUCTION

During the past few years, emotional recognition technology has emerged as an eminent and challenging technique with various applications across fields like healthcare, psychology, human-computer interaction and biomedical engineering. This helps not only to understand and detect emotions but also to evaluate physiological parameters like blood pressure or stress level and eventually diagnose psychological disorders as well as enhance user experience in intelligent environment.

The significance of emotion recognition is demonstrated by its power to pick up on subtle facial expressions and gestures that reveal underlying emotional states. However, despite its importance there is no generally accepted definition for emotion which points out its complexity and multi-dimensionality. Scholars have been researching on emotion detection intensively by investigating diverse modalities such as; facial expressions, Galvanic Skin Response (GSR), Electroencephalography (EEG) and visual scanning behavior.

Facial expression analysis among these modalities emerges as a predominant strategy towards detecting emotions. It entails several tools such as feature extraction, action unit identification and deep learning methods. Although there has been considerable progress in controlled environments, there remain challenges in emotion recognition under naturalistic conditions due to variations in facial pose, lighting, and subtle expression differences. Addressing these challenges calls for innovative approaches, together with rigorous experimentation, to improve prediction accuracy and model performance.

Literature reviews of emotion recognition methodologies, datasets, and performance metrics have indeed proven quite useful. They explored the performance of various CNN architectures, including VGG networks, ResNet, Mini-xcepetion and ensemble models, for recognition of facial expressions. Moreover, studies also investigate a number of datasets, including FER2013, AffectNet, JAFFE, CK+, and KDEF, to benchmark model performance and identify areas that call for improvements.

FER2013

FER2013, short for Facial Expression Recognition 2013, is a dataset with images of faces annotated for seven different expressions: anger, disgust, fear, happiness, sadness, surprise, and neutral. It is one of the most popular in the field of facial expression recognition.

- FER2013 features over 35,000 grayscale images of size 48x48 pixels, which is further divided into

***Corresponding Author:** Sakshi Gupta

three sets: training, validation, and test.
- FER2013 is a common dataset used for training and evaluating face emotion models, especially those relying on deep learning.

AffectNet
- AffectNet is a large-scale dataset that can be used for analyzing facial expression and is specifically designed to analyze human facial expressions.
- It contains more than 1 million facial images with labeled emotions, including basic emotions as well as complex emotional states.
- The dataset covers a wide range of ages, ethnicities, and lighting conditions, making it suitable for training robust models for recognizing facial expressions.
- AffectNet is useful for research on emotion recognition and human-computer interaction.

JAFFE (Japanese Female Facial Expression)
- The JAFFE dataset has facial images of 21 female Japanese subjects, showing six basic facial expressions: anger, disgust, fear, happiness, sadness, and surprise.
- The dataset is made up of 213 grayscale images, each with a resolution of 256x256 pixels.
- JAFFE is commonly used in facial expression recognition research at an early stage due to its relatively small size, especially for explaining issues with algorithm development.

CK+ (Extended Cohn-Kanade)
- CK+ is an extended version of the Cohn-Kanade dataset, specifically developed for facial expression analysis.
- It consists of face image sequences from 123 subjects, each displaying a range of facial expressions, including neutral, anger, contempt, disgust, fear, happiness, sadness, and surprise.
- CK+ contains both posed and spontaneous expressions, therefore making it useful to study spontaneous emotion recognition.
- It is one of the most commonly used datasets to benchmark and train systems dealing with facial expressions.

KDEF (Karolinska Directed Emotional Faces)
- The KDEF dataset features facial images of a number of actors displaying six basic emotional expressions: anger, disgust, fear, happiness, sadness, and surprise.
- The dataset consists of 4900 color images with each actor expressing every emotion multiple times under different lighting conditions and camera angles.
- The dataset has been used in many different fields like research on facial expression recognition, affective computing, and psychology.
- It provides a standardized set of facial expressions for the evaluation of the performance of algorithms used in facial expression analysis.

This paper tries to give a review of already available literature surveys on facial emotion recognition. Using the findings from such analyses, we seek to know whether there are commonalities and challenges in the field and what the future directions of research should be. By combining the above knowledge and even presenting new methodologies, this paper aims at promoting the advancement of facial emotion recognition technology and its application in various fields.

METHODS

The proposed technique for identifying emotional expressions in the (1) is Enhanced Convolutional Neural Network (ECNN). Specifically, it aims at enhancing emotion recognition via face features by recognizing those facial expressions which can elicit human responses. ECNN uses a convolutional neural network to identify seven basic emotions and examines various pre-processing techniques to figure out how they affect the performance of CNN. This paper concentrates on enhancing the facial attributes and expression based on emotion recognition with an aim of improving Emotion Detection through Face Analysis. The results show that ECNN approach has been able to achieve significantly improved accuracy compared to existing methods achieving a total accuracy rate of around 97% when implemented.

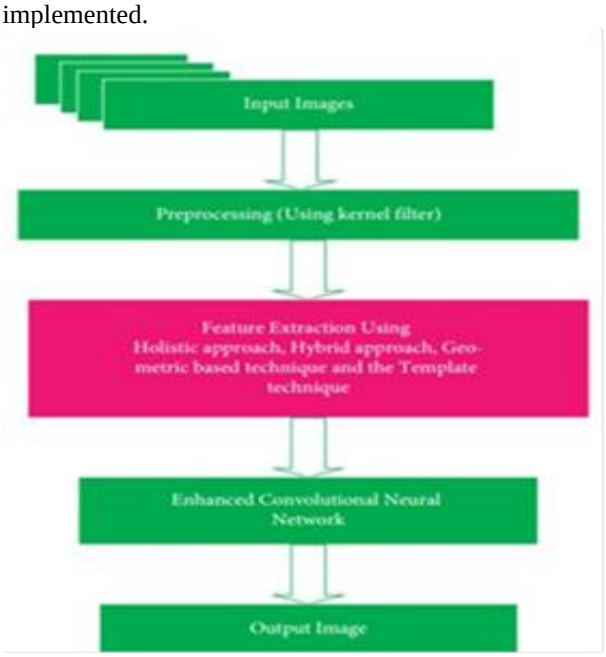

Figure 1: ECNN Model

Figure 1 gives the ECNN approach in which, the model utilizes a two-level Convolutional Neural Network (CNN) architecture for facial expression recognition (FER). The initial stage involves background removal to isolate facial expressions using standard CNN modules. Subsequently, facial features

***Corresponding Author:** Sakshi Gupta

are detected using additional filtering methods. Feature extraction techniques such as geometric, holistic, hybrid, and template-based approaches are employed to classify facial expressions accurately. Pre-processing techniques including kernel filtering for noise reduction and edge detection are applied prior to feature extraction. Finally, the CNN model processes the extracted features to classify facial expressions through layers of convolution, RELU activation, and max-pooling, ultimately providing accurate recognition results.

The proposed technique for identifying emotional expressions in (2) is the development of a Mini-Xception model based on the Xception and Convolutional Neural Network (CNN) to automatically detect and predict emotional conditions with high accuracy. Deep learning techniques are applied in the Mini-Xception model to recognize emotions, classify them, and detect them using FER-2013 dataset. Another feature of this model is that it can detect seven kinds of emotions: anger, disgust, fear, happy, sad, surprise and neutral with an accuracy of around 95.60%. The proposed structure of the model enables improved recognition capability through blending residual convolution networks with depth-wise separable convolutions.

The proposition for expression of emotions recognition in the paper (4), stays centered on employing convolutional neural networks (CNNs) for facial emotion recognition. This document also shows effectiveness of CNNs in recognizing facial expressions and highlights various methods used to accomplish FER among them being transfer learning, custom CNNs and ensemble models. Another aspect of the paper is about using binary models for FER, which is not commonly adopted in existing literature thereby extending the range of methods available for FER and setting the stage for new avenues of inquiry and exploration in emotion recognition research. Dataset modification techniques such as filtering and augmentation are studied by this work. Additionally, there is a comparison analysis made on different existing FER systems which use CNNs while introducing two new novel efficient prioritized models and also update the FER database making a good contribution towards facial emotion recognition with convolutional neural networks.

The emotion expression recognition of the proposed method in the paper(3) applies Convolutional Neural Networks (CNNs) to achieve the best single-network classification accuracy on the FER2013 dataset. This approach involves vigorous hyperparameter tuning of the VGGNet architecture and fine-tuning it by several optimization methods. The paper makes the prediction of improved FER2013 prediction using CNNs and the overall state-of-the-art accuracy achieved is 73.28% without using any extra training data. Furthermore, several saliency maps have also been constructed to better understand the performance and decision-making process of the network.

Paper (5) suggested method of recognizing emotion from documentarians can be described as a real-time prediction and recognition system by means of a Convolutional Neural Network (CNN) algorithm for recognizing facial emotions. The system is supposed to use the dataset of Facial Emotion Recognition (FER-2013), which incorporates grayscale images of dimensions 48x48 pixels for each image and contains seven distinct types of micro-expressions. This will help in face detection, facial feature extraction, and facial emotion classification by using the CNN approach. The system predicts and recognizes seven basic human expressions, which are essentially happy, sad, angry, scared, surprise, contempt, and disgust, with an average accuracy of 65.97%. The method also proposes the testing of the system's accuracy under different conditions, like face position, distance from the camera, and image rotation, to evaluate its usefulness in real-life applications. Other improvements for the system involve the use of a high-resolution camera and increasing the number of training data to enhance accuracy, especially for tests with varying distances, viewing angles, and image rotations.

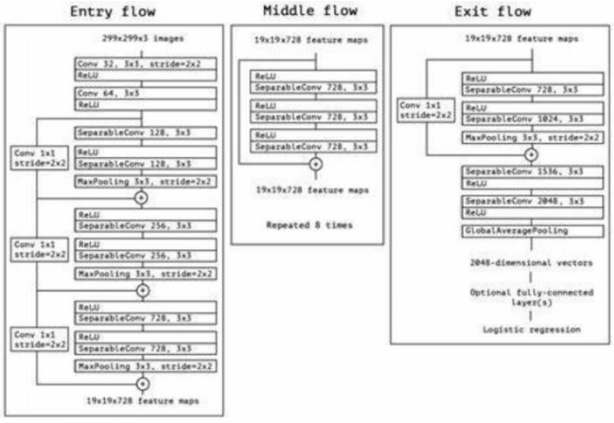

Figure 2: Mini-Xception Architecture

Figue 2 represents Mini-Xception model in which, the architecture is divided into 3 parts, the Entry, Middle and Exit flow.

Entry Flow: Performs initial feature extraction and down sampling.

- SeparableConv2D: Applies depth-wise separable convolutions to extract features efficiently.
- ReLU: Adds non-linearity with the ReLU activation function.
- Max Pooling 2D: Reduces the feature map size for faster processing.

Middle Flow: Repeats entry blocks for deeper feature learning.

***Corresponding Author:** Sakshi Gupta

Entry block:

- SeparableConv2D: Extracts features with depth-wise separable convolutions.
- ReLU: Normalize and add non-linearity.
- Residual connection: Adds the output of the first SepConv2D to the output of the second.

Exit Flow: Extracts high-level features and generates predictions.

- SeparableConv2D: Extracts final high-level features.
- GlobalAveragePooling2D: Converts feature maps into a single vector.
- Logistic Regression: It is used to label each image with corresponding emotions.

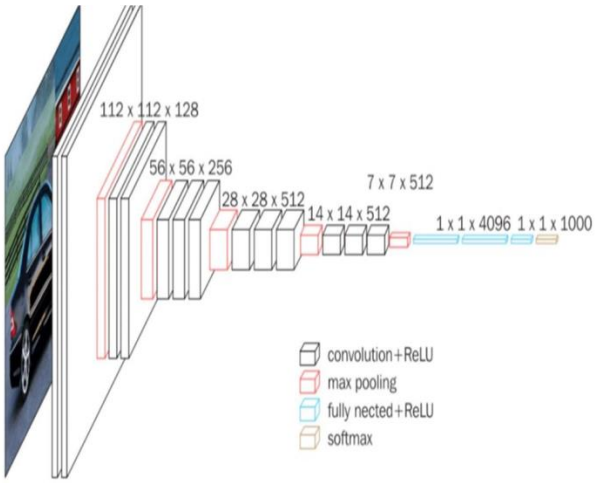

Figure 3: VGGNet Architecture

The Figure 3 depicts the VGG-16 network broken down into its various layers.

- Input Layer: This layer takes in an image of size 224x224 pixels.
- Convolutional Layers: These layers are the core building blocks of CNNs. They apply filters to the input image, extracting features like edges and shapes. VGG-16 uses multiple sets of convolutional layers, each with a small filter size (3x3 pixels) and a non-linear activation function (ReLU)
- Pooling Layers: These layers reduce the dimensionality of the data by summarizing the outputs of the convolutional layers. In the image, maxpooling is used, which down samples the data by selecting the maximum value from a sub region.
- Fully Connected Layers: These layers connect all the neurons in one layer to all the neurons in the next layer, allowing the network to learn more complex relationships between features. There are three fully-connected layers in VGG-16, each followed by a ReLU activation function.
- Output Layer: This layer uses a softmax function to classify the image into one of 1000 categories.

EVALUATION METRICS

Evaluation metrics are quantitative measures that help determine the performance of a machine learning model or system. In the image classification scenario, evaluation metrics are helpful in determining how well the model is able to classify images appropriately based on the represented emotion. Some commonly used evaluation metrics for image classification are as follows:

1. Accuracy: Accuracy measures the proportion of correctly classified images from all the images in the dataset. It is calculated as the number of correct predictions divided by the total number of predictions.
2. Precision and Recall: Precision refers to the percentage of true positives (i.e., true predictions on positive samples) among all instances predicted as positive. Recall, also known as sensitivity, refers to the percentage of true positives among all actual positive samples in the dataset. They are particularly useful in dealing with class imbalance, where one class (emotion) may predominate over others.
3. F1 Score: F1 is the harmonic mean of precision and recall. F1 plays an imperative role as it gives a balance between precision and recall for the case where there is an uneven class distribution or where the equally important false positives and false negatives are simultaneously to be identified.
4. Confusion Matrix: The confusion matrix is a table summarizing the classification model's performance by comparing the predicted label values with the actual label values. The confusion matrix can provide the classification error including false positives, false negatives, true positives, and true negatives.
5. Receiver Operating Characteristic (ROC) Curve and Area Under the Curve (AUC): The ROC curve plots the true positive rate (recall) against the false positive rate (1 - specificity) at various threshold settings. The AUC represents the area under the ROC curve, providing an aggregation value to evaluate the performance of the model across all the possible thresholds.
6. Mean Average Precision (mAP): mAP is commonly used in object detection tasks but can also be applied to image classification problems. It computes the average precision across all classes and has special relevance to many-class classification problems.
7. Cross-Entropy Loss: The cross-entropy loss (or log loss) is a measure of difference between the predicted probability distribution and the true probability distribution of the classes. A lower cross-entropy loss indicates better model performance.

RESULTS

With the most impressive accuracy rates and with emphasis on refinement forthe faces and their expressions to help in emotion recognition, the

***Corresponding Author:** Sakshi Gupta

Enhanced Convolutional Neural Network (ECNN) method, as described in(1),giving 97% accuracy using the precision and recall evaluation metric. In terms of the second approach, the Mini-Xception Model described in (2) not only achieved high accuracy but also showed the capability for real-time detection and classification of emotions. This model had 95.60% accuracy and used accuracy, precision, recall and confusion metrics for evaluation. Combining these two methodologies complements each other to offer distinctive features and qualities, adding up to provide the fundamental basis for the development of research in emotion expression recognition.

CONCLUSION

With the integration of Convolutional Neural Networks (CNNs) and deep learning techniques, there has been a substantial leap in emotion recognition. Many papers reviewed employ the Enhanced Convolutional Neural Network (ECNN) and the Mini-Xception Model to register accuracy percentages going over 95 percent. This indicates the power of deep learning in realizing fine-grained characteristics and expressions through facial features that are captured, and this will also improve the realism of emotion classification tasks. Residual connections and ensembles are some of the innovations that these architectures leverage for flexibility in identifying faces that express different emotions. Another important aspect of emotion classification models is the use of datasets such as FER-2013, offering generalization and robustness by promoting interdisciplinary collaboration. While impressive advancements are attained, new avenues of exploration include novel architectures, multi-modal techniques, and deployment in real-world scenarios. Future advancements, backed by emerging technologies, are expected to transform the very deep learning-based technology in emotion expression recognition, which would be essential for interdisciplinary collaboration.

REFERENCES

1. Kumar Arora, T., Kumar Chaubey, P., Shree Raman, M., Kumar, B., Nagesh, Y., Anjani, P. K., ... & Debtera, B. (2022). Optimal facial feature based emotional recognition using deep learning algorithm. *Computational Intelligence and Neuroscience*, *2022*(1), 8379202.
2. Fatima, S. A., Kumar, A., & Raoof, S. S. (2021). Real time emotion detection of humans using mini-Xception algorithm. In *IOP conference series: materials science and engineering* (Vol. 1042, No. 1, p. 012027). IOP Publishing.
3. Khaireddin, Y., & Chen, Z. (2021). Facial emotion recognition: State of the art performance on FER2013. *arXiv preprint arXiv:2105.03588*.
4. Białek, C., Matiolański, A., & Grega, M. (2023). An Efficient Approach to Face Emotion Recognition with Convolutional Neural Networks. *Electronics*, *12*(12), 2707.
5. Zahara, L., Musa, P., Wibowo, E. P., Karim, I., & Musa, S. B. (2020, November). The facial emotion recognition (FER-2013) dataset for prediction system of micro-expressions face using the convolutional neural network (CNN) algorithm based Raspberry Pi. In *2020 Fifth international conference on informatics and computing (ICIC)* (pp. 1-9). IEEE.

***Corresponding Author:** Sakshi Gupta