SJIF Impact Factor (2023): 8.574| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 8 | Issue: 7 | July 2023 - Peer Reviewed Journal

FACIAL EMOTIONAL RECOGNITION USING MOBILENET BASED TRANSFER LEARNING

Jenisha A[1*], Aleesha Livingston L[2]
[1*]*ME/Communication Systems, Bethlahem Institute of Engineering*
[2]*Assistant Professor/Electronics & Communication Engineering, Bethlahem Institute of Engineering*

[*]*Corresponding Author: Jenisha A*

ABSTRACT
In the real world detecting a facial emotion is challenging and complicated. To identify the subtle differences in facial expressions, Facial Emotion Recognition (FER) requires the model to learn. For image recognition tasks, a convolutional neural network (CNN) is a type of deep learning model that is commonly used. CNNs are able to learn features from images that are relevant to the task at hand, such as facial expressions. A pre-trained CNN is a CNN that has already been trained on a large dataset of images for another task, such as image classification. Pre-trained CNNs can be used to improve the performance of CNNs for other tasks, such as facial emotion recognition. The main difference between a CNN and a pre-trainedCNN is that a pre-trained CNN has already learned to extract features from images that are relevant to the task at hand. This means that a pre-trained CNN can be used to improve the performance of a CNN for the task at hand without having to train the CNN from scratch. Here we use MOBILENET as the pre-trained convolution neural network used with the help of the transfer learning technique. MOBILENET is a pre-trained CNN for FER, because it is efficient and accurate.EmoNet is a proposed mobile facial expression recognition system that utilizes the power of transfer learning and the efficiency of the MOBILENET model. The system aims to accurately classify facial expressions in real-time on mobile devices, making it accessible and user-friendly. The data is collected, pre-processed, and fed into the MOBILENET model for feature extraction. Stochastic gradient descent (SGD) is employed to train the pre-processed model, and its performance is evaluated using precision, recall, F1-measure, and accuracy metrics. Through experimental analysis and performance visualization, EmoNet demonstrates high estimation values and superior severity-level classification results compared to other models. This system offers a promising solution for efficient and accurate facial expression recognition, with potential applications in various domains, including emotion detection, human-computer interaction, and social robotics.
KEYWORDS: *CNN, EmoNet, Facial Emotional Recognition, MOBILENET, Pre-trained CNN.*

1. INTRODUCTION

Based on facial expressions,Facial Emotional Recognition (FER) is a technology that aims to detect and analyze human emotions [1]. One real-world problem with FER is its limited accuracy, especially when it comes to recognizing subtle emotional cues or expressions in individuals from diverse cultural backgrounds. For example, certain cultures may have unique facial expressions that differ from the standard dataset used for training FER models, leading to misinterpretations [2]. Challenges in FER include handling variations in lighting conditions, occlusions, and individual differences in facial features. [3]

Deep neural networks, which require large amounts of labeled data and computationally intensive training asexisting techniques for FER rely on machine learning algorithms. However, these techniques often suffer from biases, lack of generalizability, and privacy concerns associated with facial data collection [4]. Moreover, FER systems can be susceptible to adversarial attacks, where slight modifications to a facecan deceive the system into misclassifying emotions [5]. Addressing thesedrawbacks and developing more robust, culturally diverse, and privacy-respecting FER models remains an ongoing challenge.

The architecture of EmoNet model for Facial emotion recognition makes several significant contributions,

- EmoNet leverages MobileNet and transfer learning to achieve efficient and accurate facial expression recognition on mobile devices.
- Designed for mobile devices, providing accessibility and convenience for facial expression recognition tasks.
- Demonstrates high estimation values and outperforms other models in severity-level classification, ensuring superior accuracy and effectiveness.

SJIF Impact Factor (2023): 8.574| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 8 | Issue: 7 | July 2023 - Peer Reviewed Journal

- Potential applications in various domains, such as emotion detection, human-computer interaction, and social robotics, enhancing experiences and interactions in these areas.

The paper's remaining sections are as follows: Section 2 our proposed EmoNet model have been explored in facial emotion recognition, including traditional methods and deep learning approaches. Challenges remain in accurate detection and efficient deployment on resource-constrained devices. In section 3, Proposed EmoNet Model is a mobile facial expression recognition system that combines transfer learning with MobileNet. It uses MobileNet as a pre-trained CNN for feature extraction, followed by training with stochastic gradient descent (SGD). In section 4, Extensive experiments and performance visualization demonstrate EmoNet's accuracy and effectiveness in facial emotion recognition as accuracy metrics. In section 5, EmoNet offers an efficient and accurate solution for real-time facial expression recognition on mobile devices. Its transfer learning and MobileNet integration make it accessible and user-friendly, with potential applications in emotion detection, human-computer interaction, and social robotics domains.

2. RELATED WORKS

Some of the papers based on the facial emotional recognition are reviewed below,
In their work, Kim and Song [6], proposed to generate feature transformation for emotional expression representation, quantifying contrast between facial features, and recognizing emotions based on polar coordinate understanding of angle and intensity (i.e.) Arousal-Valence space.

Banskota*et al*. [7] proposed, a modified CNNEELM approach was employed to enhance accuracy and reduce processing time in facial emotion recognition during training. The system incorporated optical flow estimation for motion detection in facial expressions and extraction of peak images. Successfully recognizing six facial emotions (happy, sad, disgust, fear, surprise, and neutral) was achieved using the proposed CNNEELM model.

Siddique*et al*. [8] proposed, a standardized framework for comparing and contrasting FER models. A lightweight convolutional neural network (CNN) was trained on the AffectNet dataset, which is a diverse and extensive dataset for facial emotion recognition. The CNN was embedded within the application and demonstrated the capability of instant, real-time facial emotion recognition.
Santoso and Kusuma [9] modified, the classification layer with the SpinalNet and ProgressiveSpinalNet architectures andadopted the state-of-the-art models in ImageNetto

improve theaccuracy. The classification was performed on the FER2013 dataset, which was openly shared with the public on Kaggle.

Devaram*et al*. [10] proposed, a compact and robust service named Lightweight EMotionrecognitiON (LEMON) for Assistive Robotics. LEMON leveraged image processing, Computer Vision, and Deep Learning (DL) algorithms to effectively recognize facial expressions. The DL model employed in the research was built upon Residual Convolutional Neural Networks, which integrated a blend of Dilated and Standard Convolution Layers.

Alamgir and Alam [11] proposed,to identify and categorize facial expressions into seven distinct emotions. The collected dataset images underwent pre-processing to remove noise, followed by extraction of significant geometric and appearance-based features. From the extracted feature set, the most relevant features were carefully selected.

Kumari and Bhatia [12] proposed, deep learning-based FER tool. Initially, the obtained dataset was applied to a joint trilateral filter to remove the noise. Then, contrast-limited adaptivehistogram equalization was applied to the filtered images toimprove the visibility of images.

Alsharekh [13] proposed, a CNN model was employed as an efficient DL technique for emotion classification from facial images. The algorithm introduced an enhanced network architecture specifically tailored to handle aggregated expressions detected by the Viola Jones (VJ) face detector. Through a series of experiments, the internal architecture of the proposed model was fine-tuned to achieve optimal performance.

Vats and Chadha [14] proposed, FER framework, Swin Vision Transformers (SwinT) and squeeze and excitation block (SE) were utilized to tackle vision tasks. The approach involved incorporating an attention mechanism, SE, and SAM within a transformer model to enhance model efficiency, considering transformers typically demand extensive data.

Dubey *et a*l. [15] provided,a comprehensive assessment of the current progress in this field was conducted, analyzing both the potential benefits and challenges associated with the adoption of facial emotion recognition technology. Its rising popularity was observed in movie and music recommendation systems.

Table 1: Comparative analysis of the existing methods on facial emotional recognition

Authors	Methods Used	Advantages	Disadvantages
Kim and Song [6]	CNN	-Generating features related to emotional expression and enhances emotional representation learning.	-Depend on the quality and diversity of the training data, and its generalizability to different datasets.
Banskota*et a*l. [7]	CNNEELM	-Improved accuracy during the training session.	-may introduce errors or limitations in capturing subtle facial expression changes.
Siddique*et al*. [8]	CNN	-It provides a Standardized approach for comparing and contrasting FER models, allowing for more consistent evaluation.	-Limit the recognition accuracy compared to more complex and deep neural networks, potentially affecting.
Santoso and Kusuma [9]	VGGNet	-modification of the classification layer improved the accuracy of the classification results.	-maylimit the generalizability of the findings to other datasets or real-worldscenarios.
Devaram*et al*. [10]	DRCNN	- It provides a compact and robust solution in Assistive Robotics, enabling betterhuman-robot interaction and assistance.	-Its complexity and Resource requirements, may pose challenges in real-time implementation and deployment.
Alamgir and Alam [11]	DBRO	-The potential to accurately identify and categorize facial expressions, which can aid in various applications.	-Accurately extracting relevant features and achieving robust performance across different datasets and variations.
Kumari and Bhatia [12]	CNN	-Recognition tool can effectively reduce noise and enhance image visibility, improving the accuracy.	-Enhancement techniques may introduce artifacts or distortions in the images, potentially affecting the accuracy.
Alsharekh [13]	CNN	- It provides an efficient and accurate approach for classifying emotions from facial images.	-Limitations in accurately capturing subtle facial cues, potentially affecting the performance of the proposed algorithm.

SJIF Impact Factor (2023): 8.574| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 8 | Issue: 7 | July 2023 - Peer Reviewed Journal

Vats and Chadha [14]	SwinT	-Squeeze and excitation block improves the efficiency and effectiveness, even with limited data.	-Limit the model's ability to capture the full complexity and variability of facial expressions, potentially affecting the accuracy and generalizability.
Dubey *et al*.[15]	CNN	-Music and movie recommendation systems had become popular.	-Challenges such as accurate emotion detection, and limited availability of labeled emotional data pose difficulties in implementing system at scale.

3. PROPOSED EMONET: MOBILE FACIAL EXPRESSION RECOGNITION SYSTEM

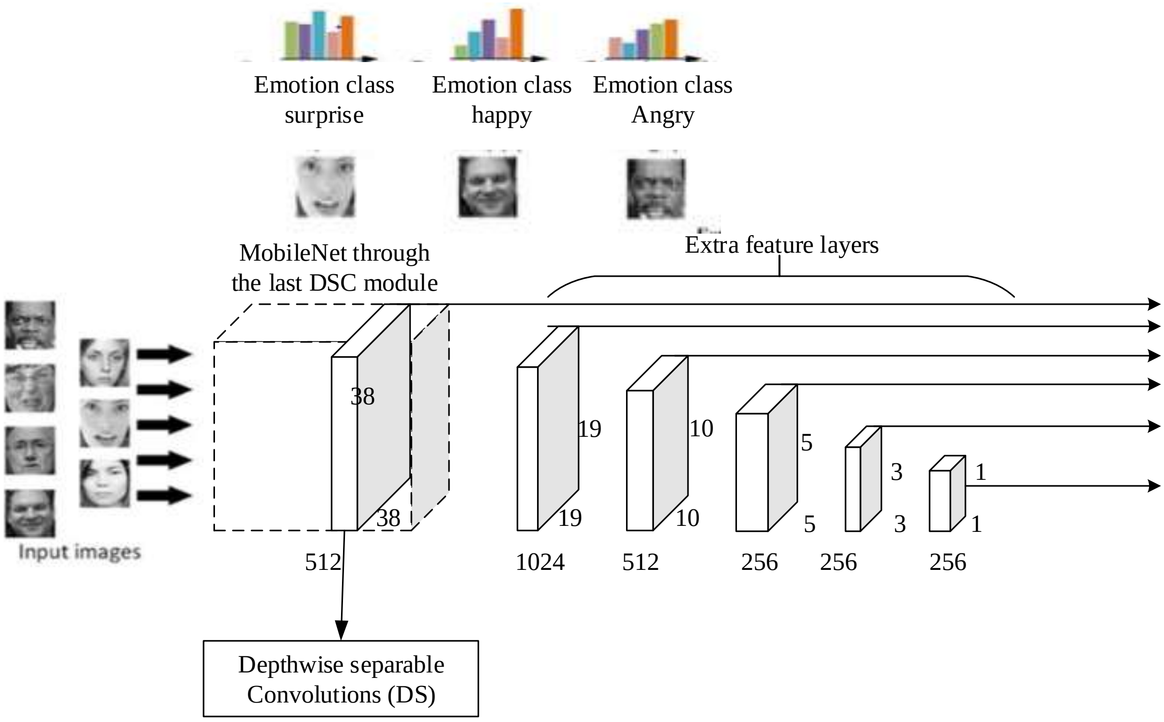

Figure 1: Architecture of EmoNet model

The challenges in Facial Emotional Recognition (FER) include the quality and diversity of training data, limitations in optical flow estimation, difficulties in real-time implementation and deployment, capturing subtle facial cues, extracting relevant features, and addressing the scarcity of labeled emotional data. Overcoming these challenges is essential for the development of accurate and reliable FER systems.In this research, the proposed system EmoNet employs MOBILENET as a pre-trained CNN for facial expression recognition. It involves data collection and pre-processing, feature extraction using the MOBILENET model, training the model with SGD, evaluating its performance, analyzing results, and comparing them with other models. The approach demonstrates high accuracy and effective classification of facial expressions.

3.1 Data Collection and Pre-processing

The facial expression recognition system collects data, which consists of images or video frames of faces with labeled expressions. The collected data is pre-processed by normalizing the pixel values, for normalizing pixel values common technique is used called

SJIF Impact Factor (2023): 8.574| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 8 | Issue: 7 | July 2023 - Peer Reviewed Journal

as min-max normalizationand also removing some noises likeGaussian, salt and pepper, speckle noise, motion blur, and illumination variations are present in the images.

normalized_value = (pixel_value - min_value) / (max_value - min_value)

3.2 Pretraining the EmoNet Model

EmoNet model refers to the process of using a pre-trained convolutional neural network (CNN) called MobileNet as the foundation for the EmoNet model. MobileNet is already trained on a large dataset for a different task, such as image classification, and has learned to extract relevant features from images. By leveraging transfer learning, EmoNet takes advantage of MobileNet's pre-learned feature extraction capabilities and fine-tunes it specifically for facial expression recognition. Already trained MobileNet CNN as a starting point and adapting it to accurately classify facial expressions. This approach saves time and computational resources by avoiding the need to train the model from scratch and enhances the performance of EmoNet for facial emotion recognition.

Mobile Net

MobileNet is a pre-trained CNN architecture designed for efficient and accurate facial expression recognition on mobile devices. It applies convolutional layers to extract features from input facial images, using depthwise separable convolution to reduce computations. Pooling layers downsample the feature maps, capturing high-level features. Fully connected layers classify the features into facial expressions, and a softmax activation produces probability scores. MobileNet's lightweight design enables real-time emotion recognition on mobile devices, making it suitable for emotion detection, human-computer interaction, and social robotics applications.

Depthwise separable convolutions

Depthwise separable convolutions in CNNs reduce computational complexity by applying separate convolutions to input channels and combining the outputs through pointwise convolutions, resulting in efficient models suitable for mobile devices. By separating spatial and channel-wise operations, depthwise separable convolutions significantly reduce parameters and computations, maintaining accuracy while enabling deployment on resource-constrained devices.

3.3 Training with Stochastic Gradient Descent

Training with stochastic gradient descent (SGD) involves dividing the pre-processed facial expression data into batches. The MobileNet model, along with additional layers, is initialized with random weights. Images are passed through the network, predictions are compared to the true labels using a loss function, and gradients are computed using backpropagation. The weights are updated using SGD to minimize the loss function. This iterative process continues for multiple epochs, gradually improving the model's performance.

Transfer Learning (TL)

Transfer learning (TL) is a technique that uses a pre-trained model's knowledge to improve performance. The pre-trained model, like MobileNet, is trained on one task (e.g., image classification) and learns to extract relevant features. These features are then applied to a different task, such as facial emotion recognition. By preserving the pre-trained model's weights and feature extraction capabilities, the model starts with a higher performance level and can converge faster. Transfer learning is particularly useful when data is limited or specialized expertise is required. It has shown effectiveness in domains like natural language processing, computer vision and audio processing, enhancing the accuracy and efficiency of deep learning models applied to new tasks.

3.4 Experimental Results

Once the model is trained, it is important to calculate its performance on a held-out test set. This step helps to determine how well the model will generalize to new and unseen data. Metrics as accuracyis computed to assess the model's performance.

Accuracy

Accuracy measures the proportion of correct predictions as true positives and true negatives out of the total instances, assessing the overall performance of a classification model like EmoNet in facial expression recognition.

$$\text{Accuracy}=\frac{TP+TN}{TI}$$

4. COMPARATIVE ANALYSIS

Table 2 indicates the performance of proposed EmoNet model with existing methods such as CNNEELM [7], and CNN [13]. The proposed EmoNet model achieved the better segmentation performance than other methods for Facial emotion recognition. Our method achieves CK^+ dataset of 97.62% and FER 2013 dataset value of 99.17%.

SJIF Impact Factor (2023): 8.574| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 8 | Issue: 7 | July 2023 - Peer Reviewed Journal

Table: 2 CK⁺ dataset and FER 2013 dataset for proposed and existing methods

Methods	Accuracy	
	CK^+ dataset (%)	FER 2013 (%)
CNNEELM [7]	96.23	98.11
CNN [13]	90.98	89.2
Proposed EmoNet	97.62	99.17

Figure: 2 Recognition Results

5. CONCLUSION

In this paper, facial emotion recognition is a challenging task that requires identifying subtle differences in expressions. Transfer learning with MobileNet, an efficient and accurate pre-trained CNN, enhances performance. EmoNet utilizes MobileNet and transfer learning for real-time, user-friendly facial expression classification on mobile devices. Using stochastic gradient descent and evaluation metrics, EmoNet achieves high estimation values and superior classification compared to other models. It offers an effective solution for precise and efficient facial expression recognition, applicable to emotion detection, human-computer interaction, and social robotics.

REFERENCES

1. *Mellouk W, Handouzi W (2020) Facial emotion recognition using deep learning: review and insights. Procedia Computer Science, 175:689-694.*
2. *Canedo D, Neves AJ (2019) Facial expression recognition using computer vision: A systematic review. Applied Sciences, 9(21):4678.*
3. *Samadiani N, Huang G, Cai B, Luo W, Chi CH, Xiang Y, He J (2019) A review on automatic facial expression recognition systems assisted by multimodal sensor data. Sensors, 19(8):1863.*
4. *Liliana DY (2019) April. Emotion recognition from facial expression using deep convolutional neural network. In Journal of physics: conference series (Vol. 1193, No. 1, p. 012004). IOP Publishing.*
5. *Lasri I, Solh AR, El Belkacemi M (2019) Facial emotion recognition of students using convolutional neural network. In 2019 third international conference on intelligent computing in data sciences (ICDS) (pp. 1-6). IEEE.*

SJIF Impact Factor (2023): 8.574| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 8 | Issue: 7 | July 2023 - Peer Reviewed Journal

6. *Kim D, Song BC (2022) Emotion-aware Multi-view Contrastive Learning for Facial Emotion Recognition. In Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XIII (pp. 178-195). Cham: Springer Nature Switzerland.*
7. *Banskota N, Alsadoon A, Prasad PWC, Dawoud A, Rashid TA, Alsadoon OH (2023) A novel enhanced convolution neural network with extreme learning machine: facial emotional recognition in psychology practices. Multimedia Tools and Applications, 82(5):6479-6503.*
8. *Siddiqui N, Reither T, Dave R, Black D, Bauer T, Hanson M (2022) A Robust Framework for Deep Learning Approaches to Facial Emotion Recognition and Evaluation. In 2022 Asia Conference on Algorithms, Computing and Machine Learning (CACML) (pp. 68-73). IEEE.*
9. *Santoso BE, Kusuma GP (2022) Facial emotion recognition on FER2013 using VGGSPINALNET. Journal of Theoretical and Applied Information Technology, 100(7):2088-2102.*
10. *Devaram RR, Beraldo G, De Benedictis R, Mongiovì M, Cesta A (2022) LEMON: a lightweight facial emotion recognitionsystem for assistive robotics based on dilated residual convolutionalneural networks. Sensors, 22(9):3366.*
11. *Alamgir FM, Alam MS (2023) An artificial intelligence driven facial emotion recognition system using hybrid deep belief rain optimization. Multimedia Tools and Applications, 82(2):2437-2464.*
12. *Kumari N, Bhatia R (2022) Efficient facial emotion recognition model using deep convolutional neural network and modified joint trilateral filter. Soft Computing, 26(16):7817-7830.*
13. *Alsharekh MF (2022) Facial Emotion Recognition in Verbal Communication Based on Deep Learning. Sensors, 22(16):6105.*
14. *Vats A, Chadha A (2023) Facial Emotion Recognition. arXivpreprint arXiv:2301.10906.*
15. *Dubey A, Shingala B, Panara JR, Desai K, MP S (2023) Digital Content Recommendation System through Facial Emotion Recognition. Int. J. Res. Appl. Sci. Eng. Technol, 11:1272-1276.*