International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 03 | March - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

A Novel Neural Network Architecture for Facial Emotion Recognition

E. Manideep, G. Bhagavadgita, K. Sri Saanvi
Department of Computer Science and Engineering
Guru Nanak Institute of Technical Campus, Telangana
Guide: **A. Sunitha** (Assistant Professor)

ABSTRACT

Facial Emotion Recognition (FER) has become an important research area in computer vision and artificial intelligence due to its wide range of applications in human–computer interaction, healthcare, e-learning, surveillance, and marketing. Traditional machine learning techniques such as Random Forest and Support Vector Machines rely heavily on handcrafted features, which limits their ability to generalize across complex and diverse datasets. To address these limitations, this paper proposes a deep learning-based FER framework using the MobileNetV2 architecture. MobileNetV2 is a lightweight convolutional neural network designed for mobile and embedded devices while maintaining high accuracy. The proposed model utilizes transfer learning to classify six fundamental human emotions: happiness, sadness, anger, fear, surprise, and disgust.

Index Terms: Facial Emotion Recognition, Deep Learning, MobileNetV2, CNN, Transfer Learning, Computer Vision.

1. INTRODUCTION

Human emotions are a critical aspect of communication and interaction. Facial expressions provide one of the most important cues for understanding emotions. Automatic Facial Emotion Recognition (FER) systems aim to detect human emotions from facial images using machine learning and computer vision techniques. With the rapid development of deep learning, Convolutional Neural Networks (CNNs) have significantly improved the performance of FER systems. CNN models automatically learn hierarchical feature representations from raw images, eliminating the need for manual feature extraction. Among various CNN architectures, MobileNetV2 has gained popularity due to its lightweight design and computational efficiency.

This research proposes a MobileNetV2-based FER system that leverages transfer learning to classify six basic emotions from facial images. The system aims to provide high accuracy while maintaining efficiency for deployment on real-time and mobile platforms.

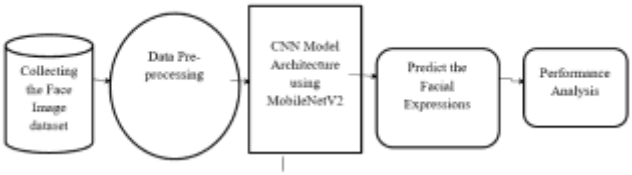
2. LITERATURE REVIEW

Carter (2023) investigated deep CNN architectures for FER and demonstrated that CNN models outperform traditional machine learning classifiers due to their ability to learn hierarchical spatial features from facial images. Mehta (2024) explored lightweight CNN architectures such as MobileNet and ShuffleNet for real-time FER applications on mobile devices. Ivanova (2023) examined transfer learning with MobileNet variants for facial emotion classification. The study demonstrated that MobileNetV2 provides better accuracy and efficiency compared to MobileNetV1 due to improved feature representation and optimized architecture. Kim (2024) proposed a hybrid deep learning model combining CNN and LSTM networks to capture both spatial and temporal features in video-based emotion recognition systems. Rossi (2024) enhanced MobileNetV2 with attention mechanisms to improve the detection of subtle facial expressions. The study reported improved classification accuracy across several benchmark datasets.

3. SYSTEM ARCHITECTURE

The system architecture of the proposed Facial Emotion Recognition (FER) system is designed to detect and classify human emotions from facial images using a deep learning model. The process begins with collecting facial images from datasets or real-time sources such as cameras. These images represent different emotional expressions including happiness, sadness, anger, fear, surprise, and disgust, which serve as input for the system. After acquiring the images, a preprocessing stage is applied to improve the quality and consistency of the data. In this stage, images are resized to a fixed dimension suitable for the neural network. Pixel values are normalized to maintain uniformity across the dataset. Additionally, data augmentation techniques such as rotation, flipping, and scaling are used to increase dataset diversity and help the model handle variations in lighting conditions, facial orientation, and background noise. Once preprocessing is completed, the images are passed to the feature extraction stage, where the MobileNetV2 architecture is used as the main deep learning model.

MobileNetV2 is a lightweight convolutional neural network that efficiently extracts important facial features. It uses depthwise separable convolutions and inverted residual blocks to reduce computational complexity while maintaining good accuracy. The model learns meaningful facial patterns such as eye movement, eyebrow position, and mouth shape that are useful for identifying emotions.



The system also uses transfer learning, where MobileNetV2 is initialized with pre-trained weights from large datasets like ImageNet. The final layers of the network are then fine-tuned for emotion classification. This approach reduces training time and improves performance when working with limited datasets.

Finally, the extracted features are passed to a classification layer that predicts the emotion category. A softmax function generates probability values for each emotion class, and the emotion with the highest probability is selected as the final output. The predicted emotion can then be used in applications such as human–computer interaction, healthcare monitoring, education systems, and real-time emotion analysis.

4. PROPOSED SYSTEM

The proposed Facial Emotion Recognition (FER) system utilizes the MobileNetV2 deep learning architecture to efficiently classify human emotions from facial images. The system begins with dataset collection, where facial image datasets such as FER2013 and CK+ are used to train the model. These datasets contain numerous facial images labeled with different emotional expressions including happiness, sadness, anger, fear, surprise, and disgust. These labeled images serve as the primary data for training and evaluating the emotion recognition model. After collecting the dataset, the system performs data preprocessing to improve the quality and consistency of the input images. During this stage, the images are resized to a fixed dimension so that they match the input requirements of the neural network. Pixel values are normalized to ensure uniformity across the dataset. In some cases, grayscale conversion is applied to simplify the image representation. Additionally, data augmentation techniques such as rotation, flipping, and scaling are used to increase dataset diversity and help the model learn robust features. These preprocessing steps improve the model's ability to generalize and perform well on unseen data. Once preprocessing is completed, the images are passed to the feature extraction stage, where the MobileNetV2 architecture automatically extracts hierarchical features from the facial images. MobileNetV2 is designed with inverted residual blocks and linear bottlenecks, which allow the network to learn complex facial patterns while maintaining computational efficiency. These features capture important facial characteristics such as eye movement, eyebrow position, and mouth shape, which are essential for identifying different emotional expressions. The system also incorporates transfer learning, where the MobileNetV2 model is initialized with pretrained weights obtained from large-scale datasets such as ImageNet. Instead of training the model entirely from scratch, the final classification layer is modified and replaced with a fully connected layer that outputs six emotion classes. This approach reduces training time and improves the overall performance of the model, especially when working with limited datasets. Finally, the trained model performs emotion classification, where the extracted features are analyzed to determine the emotional state represented in the facial image. The system predicts one of the six emotional categories—happiness, sadness, anger, fear, surprise, or disgust—and produces the corresponding output. This classification process enables the system to accurately recognize human emotions from facial expressions and can be applied in various real-world applications such as human–computer interaction, healthcare monitoring, and intelligent systems.

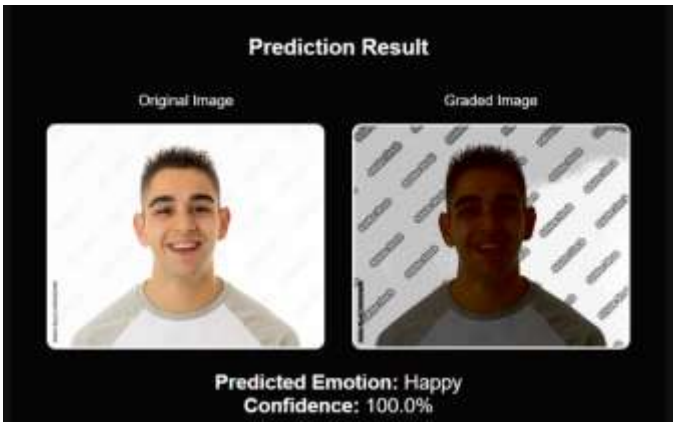
5. METHODOLOGY

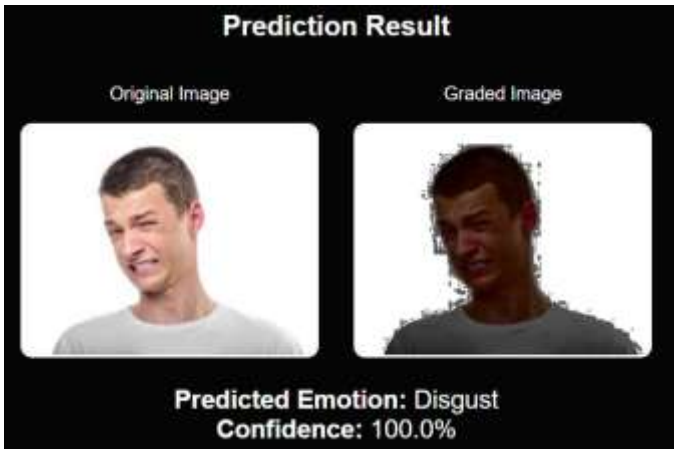
The proposed methodology for the Facial Emotion Recognition (FER) system consists of several important modules that work together to develop and evaluate the deep learning model. The first stage involves amassing the dataset, where a large collection of facial images representing different emotional expressions is gathered and properly labeled. These images serve as the training and testing data for the emotion recognition model. After collecting the dataset, the next step is transforming the data, where various preprocessing techniques such as normalization, data augmentation, and feature scaling are applied to improve the quality and diversity of the dataset. These preprocessing steps help the model learn more effectively and improve its overall performance. Once the data is prepared, the system moves to the execution stage, where the deep learning model is implemented using

Python programming language with popular frameworks such as TensorFlow or PyTorch. The model is trained using the prepared dataset so that it can learn the patterns associated with different facial emotions. After training the model, the next stage focuses on refining the accuracy of the system. Techniques such as hyperparameter tuning, dropout regularization, and optimization algorithms are applied to enhance the model's performance and prevent overfitting. Following the training and optimization stages, the system proceeds to assessing the results, where the performance of the model is evaluated using standard evaluation metrics such as accuracy, precision, recall, and F1-score. These metrics help measure how effectively the system can recognize and classify different emotional expressions. Finally, the system moves to the delivering insights stage, where the trained model can be deployed for practical use. The developed FER system can be used for real-time emotion recognition by analyzing facial images or live video streams, enabling applications in areas such as human–computer interaction, healthcare monitoring, and intelligent systems.

6. RESULTS AND DISCUSSION

The proposed MobileNetV2-based Facial Emotion Recognition (FER) system demonstrates improved performance compared with traditional machine learning approaches. One of the key advantages of the system is its ability to automatically extract meaningful features from facial images, which significantly improves classification accuracy without relying on manually designed features. The use of transfer learning further enhances the model's efficiency by reducing training time and improving its ability to generalize across different datasets. In addition, the lightweight architecture of MobileNetV2 enables the system to be deployed on mobile and embedded devices while maintaining computational efficiency. The model also shows strong robustness against variations in lighting conditions and facial orientations, which are common challenges in real-world emotion recognition tasks. Performance evaluation using confusion matrices and accuracy metrics indicates that the system can reliably classify different emotional expressions across multiple datasets.





7. CONCLUSION

This research presented a MobileNetV2-based Facial Emotion Recognition system designed to improve accuracy while maintaining computational efficiency. Unlike traditional machine learning approaches that rely on handcrafted features, the proposed model automatically learns deep features from facial images.

The integration of transfer learning significantly improves performance on limited datasets. Additionally, the lightweight architecture of MobileNetV2 allows deployment on mobile and embedded devices for real-time emotion recognition.

8. FUTURE WORK

Future enhancements of the proposed Facial Emotion Recognition (FER) system can focus on improving accuracy, functionality, and real-world applicability. One possible direction is the development of multimodal emotion recognition systems that combine facial expressions with other signals such as speech, text, and physiological data to achieve more accurate emotion detection. Another potential improvement involves exploring transformer-based deep learning architectures, which have shown promising performance in various computer vision tasks and may further enhance emotion recognition capabilities. Additionally, the integration of FER systems with augmented reality (AR) and virtual reality (VR) technologies could enable more interactive and immersive human–computer interaction experiences. Future work may also focus on implementing real-time emotion recognition on low-power devices, allowing the system to operate efficiently on smartphones and embedded platforms. Furthermore, incorporating privacy-preserving techniques such as federated learning can help protect user data by enabling model training without sharing sensitive facial images, thereby ensuring better security and user privacy.

REFERENCES

[1] A. Hassouneh et al. (2020) developed a real-time emotion recognition system that combines facial expression analysis with Electroencephalogram (EEG) signals. Their study demonstrates that integrating physiological signals with facial features can significantly improve the accuracy and reliability of emotion detection systems, particularly in applications such as healthcare monitoring and human–computer interaction.

[2] B. Schuller et al. (2003) conducted research on speech emotion recognition using machine learning techniques. Their work focuses on identifying emotions through vocal characteristics such as tone, pitch, and speech patterns. This research laid the foundation for multimodal emotion recognition systems that combine speech with other modalities like facial expressions.

[3] N. Jain et al. (2018) proposed a hybrid deep neural network approach for facial emotion recognition. Their model integrates different deep learning techniques, including convolutional neural networks, to enhance the accuracy and robustness of emotion classification systems.

[4] M. Murugappan and A. Mutawa (2021) explored facial geometric feature extraction methods for emotion classification. Their study analyzes facial landmarks such as the eyes, eyebrows, and mouth to detect subtle variations in facial expressions, demonstrating the importance of geometric features in emotion recognition.

[5] Y. Khaireddin and Z. Chen (2021) investigated the performance of deep learning models for facial emotion recognition using the FER2013 dataset. Their work provides a comprehensive analysis of different neural network architectures and highlights state-of-the-art performance in emotion classification tasks.