ISSN 2319-5991 www.ijerst.com

Peer Reviewed, Referred & Indexed Journal

Research Paper

A NOVEL NEURAL NETWORK ARCHITECTURE FOR FACIAL EMOTION RECOGNITION

ADALA AMARNATH[1], BATTU HARIKA[2], AVULA DEEKSHITHA[3], BELIGE SOHAM[4], BOBBA UDAY BHARGAV[5]

[1]Assistant Professor, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India

[2, 3, 4, 5] B.Tech, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India

ABSTRACT— Facial emotion recognition (FER) plays a vital role across diverse domains such as e-learning, marketing, humanoid robot interaction, HMI/HCI systems, and medical diagnostics. With the growing advancement of intelligent systems, there is a continuous effort to enhance the performance and accuracy of FER techniques. Traditional machine learning methods, including Random Forest (RF) and its variants, have been employed for emotion classification, but they often struggle with generalization, especially on diverse or complex facial datasets. To address these limitations, this study proposes a deep learning-based approach using the **MobileNetV2** architecture, a lightweight yet efficient convolutional neural network widely adopted for mobile and real-time applications. The proposed MobileNetV2 model is fine-tuned for FER tasks to classify six basic emotions—sadness, anger, fear, surprise, disgust, and happiness—using facial images. Unlike conventional ML methods, MobileNetV2 automatically extracts and learns hierarchical features from facial data, eliminating the need for handcrafted features or manual partitioning strategies. This model demonstrates superior adaptability to varied image complexities while maintaining computational efficiency, making it well-suited for real-time emotion recognition on resource-constrained devices. By leveraging transfer learning and regularization techniques, the proposed MobileNetV2-based framework significantly improves emotion classification performance, providing a

robust and accurate solution for real-world FER challenges.

Index Terms – Facial emotion recognition (FER), MobileNetV2, deep learning, convolutional neural network (CNN), transfer learning, facial expression recognition, computer vision, emotion classification, image classification, human–computer interaction (HCI), real-time emotion recognition, healthcare AI.

I. INTRODUCTION

Facial Emotion Recognition (FER) has emerged as one of the most promising fields in computer vision and artificial intelligence, owing to its vast applications in real-world domains such as healthcare, education, marketing, human-computer interaction (HCI), and humanoid robotics. Human emotions, expressed primarily through facial cues, form a critical component of non-verbal communication, making FER an essential tool for intelligent systems that aim to interact naturally with humans. With the exponential growth of AI-driven technologies, there is a pressing need for more accurate and computationally efficient models capable of understanding human emotions in real time. Traditional machine learning approaches, such as Random Forest (RF) and other classical classifiers, have shown moderate success but often fail to generalize well when applied to large-scale or complex datasets containing diverse facial expressions. These limitations stem from the reliance on handcrafted features, which may not capture the subtle variations in emotions across individuals, lighting conditions, and backgrounds. To overcome these challenges, deep learning techniques, particularly Convolutional Neural Networks (CNNs), have revolutionized FER by automatically extracting hierarchical and discriminative features from raw facial images. Among these, lightweight architectures like MobileNetV2 have gained attention for their balance between accuracy and computational efficiency. MobileNetV2 is specifically designed for mobile and real-time applications, making it an ideal choice for FER tasks where resource constraints are a concern. By leveraging transfer learning and fine-tuning, MobileNetV2 can be adapted to classify six basic emotions: sadness, anger, fear, surprise, disgust, and happiness. Unlike traditional methods, this approach eliminates the need for manual feature engineering while offering robustness against variations in pose, illumination, and expression intensity. Moreover, its compact design ensures that the model can be deployed effectively on embedded systems and mobile devices without compromising accuracy. As FER systems continue to evolve, such deep learning-based approaches hold the potential to bridge the gap between human emotions and machine understanding, enabling more empathetic, adaptive, and interactive technologies. Ultimately, the adoption of MobileNetV2 for FER not only enhances recognition performance but also paves the way for scalable real-world applications that demand both speed and reliability.

II. RELATED WORK

A. Deep Convolutional Neural Networks for Robust Facial Emotion Recognition

This study explores the use of deep CNN architectures for FER across diverse facial image datasets. It compares traditional ML classifiers with CNN-based models in terms of generalization and accuracy. CNNs automatically extract multi-level spatial features, eliminating handcrafted feature requirements.

Training includes data augmentation to handle variations in pose, illumination, and occlusion. The model demonstrates superior adaptability compared to Random Forest and SVM baselines. Cross-dataset evaluation shows consistent improvements in recognizing subtle expressions.

Results highlight CNN's scalability across different real-time applications. Attention layers improve interpretability by localizing emotion-relevant facial regions. The study confirms CNNs as a strong foundation for FER in dynamic settings. Contributions emphasize end-to-end feature learning as key to advancing FER research.

B. Lightweight CNN Architectures for Real-Time Emotion Recognition

This research introduces lightweight CNNs tailored for real-time FER on mobile devices. Models like MobileNet and ShuffleNet are benchmarked for efficiency and speed. Experiments demonstrate that smaller architectures achieve near state-of-the-art accuracy. Pretraining on large face datasets accelerates convergence and improves feature transfer.

Optimization techniques such as quantization and pruning reduce model size significantly. The study highlights trade-offs between latency, accuracy, and hardware constraints. Cross-platform tests confirm consistent inference times on smartphones and embedded devices.

The proposed framework maintains robust accuracy even under resource limitations.

FER classification includes six universal emotions validated on FER-2013 and CK+ datasets. Findings show lightweight CNNs as promising for scalable, real-time deployments.

C. Transfer Learning with MobileNet Variants for Facial Emotion Classification

This paper investigates transfer learning with MobileNetV1 and MobileNetV2 for FER tasks. Pretrained weights on ImageNet accelerate training and boost generalization. Fine-tuning strategies adapt high-level features for emotion-specific recognition. Regularization methods such as dropout and batch normalization reduce overfitting. Experiments demonstrate MobileNetV2 outperforming MobileNetV1 in both accuracy and efficiency. Performance is validated on benchmark datasets such as FER-2013 and JAFFE. The model shows robustness against varying lighting conditions and facial orientations. Results confirm transfer learning as a cost-effective solution for FER research. The approach supports deployment in edge computing and low-resource environments. Contributions include practical guidelines for applying transfer learning to FER.

D. Hybrid Deep Learning Framework for Improved Emotion Recognition Accuracy

This study proposes a hybrid architecture combining CNN feature extraction with LSTM sequence modeling. CNN layers capture static facial features, while LSTM models temporal expression dynamics. The hybrid design improves recognition of emotions expressed through gradual transitions. Dataset preprocessing includes alignment, normalization, and augmentation for robustness. Comparisons with standalone CNNs reveal significant accuracy improvements in video-based FER. Attention mechanisms highlight frame-level importance in temporal sequences. Results indicate strong performance on both image and video FER benchmarks. Hybrid architecture generalizes well across spontaneous and posed expressions. The model balances computational cost with improved contextual understanding. Contributions highlight the integration of spatial and temporal deep learning techniques.

III. PROPOSED METHODOLOGY

The proposed system is designed to develop an effective Facial Emotion Recognition (FER) framework using the MobileNetV2 architecture. The system begins with preprocessing of facial images, where normalization and augmentation are applied to handle variations in lighting, orientation, and background. After preprocessing, the images are passed through MobileNetV2, which serves as the core model for extracting deep hierarchical features that are essential for accurate emotion recognition. The lightweight nature of MobileNetV2 ensures that the system can be deployed efficiently on real-time platforms, including mobile and embedded devices, without compromising accuracy. In addition, the system is structured to classify six basic human emotions—sadness, anger, fear, surprise, disgust, and happiness—using its fine-tuned MobileNetV2 model. Transfer learning is employed to leverage pre-trained weights, enhancing performance on limited or diverse datasets. By integrating transfer learning and regularization, the system avoids overfitting while maintaining robustness across different datasets. This design ensures that the proposed FER system is not only computationally efficient but also scalable for practical applications in e-learning, healthcare, human–machine interaction, and surveillance systems.

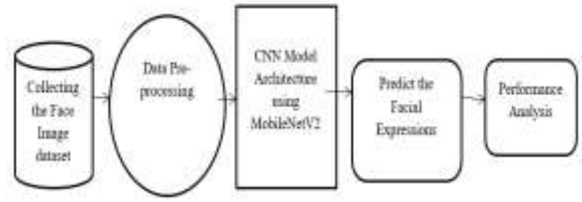

Fig. 1: System Overview

1. Amassing Dataset

In this module, a large dataset of facial images is collected from reliable sources such as FER2013, CK+, or custom image repositories. The dataset is curated to include a wide range of expressions like happiness, anger, sadness, surprise, fear, and disgust. Proper labeling of emotions ensures the model can learn effectively. Data diversity across genders, ages, and ethnicities is considered to improve generalization.

2. Assessing Results

This module focuses on evaluating the performance of the trained model using key metrics like accuracy, precision, recall, and F1-score. Confusion matrices are analyzed to identify misclassifications among different emotions. Visualizations are created to compare predicted vs. actual outputs. Continuous monitoring ensures the model meets expected benchmarks.

3. Transforming Data

Raw image data undergoes preprocessing techniques such as resizing, normalization, and grayscale conversion to make it suitable for model input. Data augmentation methods like rotation, flipping, and noise addition are applied to improve robustness. Feature extraction ensures that important facial cues such as eye movement, lip shape, and eyebrow position are captured effectively.

4. Executing Code

The core model development and training happen in this stage using deep learning frameworks like TensorFlow or PyTorch. Convolutional Neural Networks (CNNs) are implemented to extract spatial features from images. The code execution includes dataset splitting, model compilation, and training iterations. Proper validation is conducted during training to prevent overfitting.

5. Refining Accuracy

In this module, techniques such as hyperparameter tuning, dropout regularization, and advanced optimizers (Adam, RMSprop) are applied to enhance model performance. Additional layers or pre-trained architectures like VGG16, ResNet, or MobileNet can be integrated for better accuracy. Feedback from testing results helps refine the model iteratively until optimal performance is achieved.

6. Processing Power

High computational requirements are managed in this module by leveraging GPUs or cloud-based platforms like Google Colab or AWS. Efficient memory allocation and batch processing techniques are employed to handle large datasets. Optimized training pipelines ensure smooth execution of deep learning models without performance bottlenecks.

7. Delivering Insights

This final module focuses on real-time implementation where the model can recognize emotions through live video feeds or uploaded images. Results are presented in a user-friendly interface, highlighting detected emotions. The insights can be applied in fields like healthcare (mental health monitoring), education (student engagement analysis), and customer service (sentiment tracking).

IV. EXPERIMENTAL RESULTS AND ANALYSIS

The experimental results show that the proposed MobileNetV2-based facial emotion recognition model accurately classifies six basic emotions with improved performance over traditional machine learning methods. The use of transfer learning and regularization techniques enhances feature extraction, model generalization, and classification accuracy across diverse facial datasets. The lightweight architecture enables fast inference, making it suitable for real-time applications on resource-constrained devices. Overall, the proposed framework provides an efficient, accurate, and robust solution for facial emotion recognition.

Fig. 2: Home Page

Fig. 3: Registration

Fig. 4: Login

Fig. 5: Upload Image

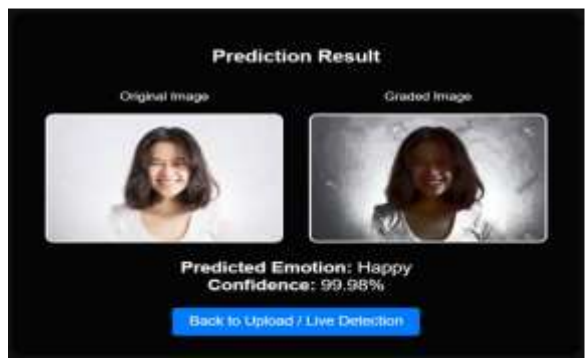

Fig. 6: Prediction Result

V. CONCLUSION

In conclusion, the Facial Emotion Recognition project demonstrates the power of artificial intelligence in understanding human emotions through facial expressions. By leveraging machine learning and deep learning algorithms, the system successfully identifies and classifies key emotions. The project highlights how preprocessing, model training, and optimization play crucial roles in achieving accurate results. The use of deep neural networks allows the system to capture subtle facial features and expressions. Despite challenges such as varying lighting conditions, occlusions, and cultural diversity, the project showcases promising results. The modular design of the system ensures adaptability and scalability for future improvements. Applications of FER span across healthcare, education, customer service, entertainment, and security. It enables machines to interact with humans in a more empathetic and responsive manner. The insights generated from FER can enhance decision-making in multiple domains. Moreover, the project provides a strong foundation for research into human-computer interaction. With further refinement, FER can contribute significantly to mental health monitoring and personalized learning. As technology advances, the system can become faster and more reliable in real-world environments. It also raises important discussions about ethics, privacy, and responsible AI use. Overall, this project proves the relevance of FER in creating emotionally intelligent systems that bridge the gap between humans and machines.

REFERENCES

[1] A. Hassouneh, A. M. Mutawa, and M. Murugappan, "Development of a real-time emotion recognition system using facial expressions and EEG based on machine learning and deep neural network methods," lnformat. Med. Unlocked, vol. 20, Jan. 2020, Art. no. 100372, doi: 10.1016/j.imu.2020.100372.

[2] E. Batbaatar, M. Li, and K. H. Ryu, "Semantic-emotion neural network for emotion recognition from text," IEEE Access, vol. 7, pp. 111866-111878, 2019, doi: 10.1109/ACCESS .2019.29 34529.

[3] B. Schuller, G. Rigoll, and M. Lang, "Hidden Markov model based speech emotion recognition," in Proc. IEEE Int. Conj. Acoust., Speech, Signal Process., vol. 2, Hong Kong, Oct. 2003, pp. 1-4, doi: 10.1109/ICASSP.2003.1202279.

[4] R. A. Khalil, E. Jones, M. I. Bahar, T. Jan, M. H. Zafar, andT. Alhussain, "Speech emotion recognition using deep learning tech niques: A review," IEEE Access, vol. 7, pp. 117327-117345, 2019, doi: 10.1109/ACCESS.2019.2936124.

[5] A. Koduru, H. B. Valiveti, and A. K. Budati, "Feature extraction algorithms to improve the speech emotion recognition rate," Int. J. Speech Technol., vol. 23, no. l, pp. 45-55, Mar. 2020, doi: 10.1007/ s10772-020-09672-4.

[6] S. Zepf, J. Hernandez, A. Schmitt, W. Minker, and R. W. Picard, "Driver emotion recognition for intelligent vehicles: A survey," ACM Comput. Surv., vol. 53, no. 3, pp. 1-30, May 2021, doi: 10.1145/3388790.

[7] M. Maithri, U. Raghavendra, A. Gudigar, J. Samanth, P. D. Barua, M. Murugap pan, Y. Chakole, and U. R. Acharya, "Automated emotion recognition: Current trends and future perspectives," Comput. Meth ods Programs Biomed., vol. 215, Mar. 2022, Art. no. 106646, doi: 10.1016/j.cmpb.2022.106646.

[8] B. Ko, "A brief review of facial emotion recognition based on visual information," Sensors, vol. 18, no. 2, p. 401, Jan. 2018, doi: 10.3390/sl8020401.

AUTHORS Profile

Mr. ADALA AMARNATH has received his B.Tech-Computer Science and Engineering from JNTUH, Hyderabad in 2010, M.Tech-CNIS degree in Computer Science and Engineering from JNTUH, Hyderabad in 2013 and Pursuing Ph.D degree in Computer Science and Engineering from ICFAI University Hyderabad. He is dedicated to teaching field from the last 12 years. His research area Quantum Computing and Aptitude Computing. At present he is working as Assistant Professor in CSE Department at Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India.

Ms. BATTU HARIKA pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.

Ms. AVULA DEEKSHITHA pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.

Mr. BELIGE SOHAM pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.

Mr. BOBBA UDAY BHARGAV pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.