International Journal of Mathematics and Computer Research

ISSN: 2320-7167

Volume 14 Special Issue 03 March 2026, Page no. – 51-57

Impact Factor: 8.751

DOI: 10.47191/ijmcr/v14iSPC3.11

Real-Time Facial Emotion Detection Using Deep Learning and AI

Padmaja Kadam, Supriya Kamareddy, Rutuja Pitrubhakta

Department of Computer Science, Dr. D. Y. Patil, Arts, Commerce & Science College, Pimpri, Pune, Maharashtra, India

ARTICLE INFO

Published Online:
14 March 2026

ABSTRACT

Facial emotion recognition has become an essential component of intelligent systems aiming to bridge the gap between human emotions and machine understanding. This project, "Real-Time Facial Emotion Detection Using Deep Learning and AI," presents an efficient approach to identifying and classifying human emotions from facial expressions in real time. The proposed system utilizes deep learning techniques, particularly Convolutional Neural Networks (CNNs), to automatically extract discriminative features from facial images and classify them into basic emotion categories such as happiness, sadness, anger, fear, surprise, disgust, and neutrality. The model is trained on publicly available facial emotion datasets to ensure high accuracy and robustness across diverse facial features, lighting conditions, and orientations. Using real-time video input from a webcam or camera feed, the system performs face detection, preprocessing, and emotion classification with minimal latency. The integration of AI-based deep learning models enables adaptive learning, improved generalization, and enhanced performance compared to traditional machine learning methods. Experimental results demonstrate that the proposed system achieves high accuracy and responsiveness, making it suitable for real-world applications such as human–computer interaction, virtual assistants, mental health assessment, and smart surveillance. The project showcases how AI and deep learning can be effectively combined to build emotionally intelligent systems capable of understanding and responding to human affective states in real time.

Corresponding Author:
Padmaja Kadam

KEYWORDS: Deep Learning, Emotion Detection, Artificial Intelligence (AI), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Dual-Tree Complex CNN (DTSCNN), Resnet-50, Facial Expression Recognition, Real-Time Processing, Computer Vision, Image Classification, Feature Extraction, Transfer Learning, FER Dataset, Human–Computer Interaction (HCI).

INTRODUCTION

Human emotions play a crucial role in communication, social interaction, and decision-making. With the rapid advancement of Artificial Intelligence (AI) and computer vision, recognizing emotions through facial expressions has become an emerging field of research known as Facial Emotion Recognition (FER). Facial expressions provide valuable non-verbal cues that reflect an individual's internal emotional state, making emotion recognition systems essential for applications such as human–computer interaction, mental health monitoring, intelligent tutoring systems, and security surveillance. Traditional methods of emotion recognition relied heavily on handcrafted features and conventional machine learning techniques, which often lacked robustness and accuracy in real-world environments. However, with the evolution of Deep Learning, particularly Convolutional Neural Networks (CNNs), significant progress has been made in automatically extracting high-level features from facial images. These advancements enable systems to achieve superior performance in identifying subtle emotional cues, even under variations in lighting, pose, and facial orientation. This project, titled "Real-Time Facial Emotion Detection Using Deep Learning and AI," focuses on developing an efficient and accurate model

Capable of recognizing human emotions from live video feeds or images in real time[1]. The system employs deep neural networks to classify facial expressions into predefined emotion categories such as happiness, sadness, anger, surprise, fear, disgust, and neutrality. By leveraging AI-driven algorithms, the model can adapt to diverse facial datasets and deliver fast, reliable predictions suitable for practical deployment. The ultimate goal of this research is to create a robust and scalable emotion recognition framework

that can enhance human–computer interaction and contribute to applications in education, healthcare, and behavioral analysis. Through real-time detection and interpretation of emotions, machines can better understand and respond to human needs, bridging the gap between technology and emotional intelligence [1]. In the modern era of artificial intelligence, understanding human emotions through computational systems has become a significant challenge and opportunity. Facial Emotion Recognition (FER) is one of the most powerful tools that bridge the gap between human affective behavior and intelligent systems. "Emotion Sense" is a Deep Learning–based Facial Emotion Recognition (FER) system designed to analyse human facial expressions in real time and classify them into categories such as happiness, sadness, anger, surprise, fear, disgust. This system uses Convolutional Neural Networks (CNNs) and computer vision techniques to process live video input and predict emotions accurately. The integration of FER into real-world applications such as healthcare, security surveillance, eLearning, and human–computer interaction provides valuable insights into emotional states, enabling more natural and adaptive systems [2].

LITERATURE REVIEW

Over the past ten years, many studies have used deep learning for facial emotion recognition (FER). Mollahosseini et al. (2017) created AffectNet, a large dataset that helped improve the accuracy of emotion recognition. Goodfellow et al. (2013) showed that deep neural networks could identify facial emotions, though their models needed a lot of computing power. Later, Zhang et al. (2021) used ResNet and VGG models to improve classification accuracy, but their systems were slow and hard to interpret[4].

Recently, researchers have focused on hybrid models and attention mechanisms to make emotion recognition more accurate. For example, a 2024 study used Coordinate Attention with MobileNetV3 to achieve better results. Some studies also used multimodal systems that combined facial expressions, voice, and body signals with transformer-based models to work better in cases of occlusion or different face angles. However, most of these systems are still too complex or too slow for real-time use. This shows the need for an easy-to-understand, fast, and accurate FER system that works well in different conditions. Emotion Sense aims to fill this gap by combining deep learning, explainable AI (XAI), and attention mechanisms to create a more practical and reliable emotion recognition tool[3].

RELATED WORK

Human Facial Expressions

The universality of facial expressions and body language is a key feature of human interaction. Charles Darwin already published on globally common facial expressions in the nineteenth century, which play an important role in nonverbal communication. In 1971, Ekman and Friesen declared facial behaviours to be correlated uniformly with specific emotions. Apparently, humans, but also animals, produce specific muscle movements that belong to a certain mental state. People interested in research on emotion classification via speech recognition are referred to Nicholson et al. Facial expressions are considered one of the most reliable indicators of an individual's emotional state, as they often occur subconsciously and spontaneously. The six universally recognized emotions—happiness, sadness, anger, fear, surprise, and disgust serve as the foundation for most emotion recognition studies. Understanding these expressions helps bridge communication gaps between humans and machines, enabling computers to interpret emotional cues more effectively. In recent years, advancements in computer vision and deep learning have significantly improved the accuracy of facial emotion analysis. As a result, emotion recognition has become a vital component in developing empathetic AI systems and enhancing human–computer interaction [7].

Figure 1: Human Facial Expressions

Image Classification Techniques

Facial emotion detection is a specialized case of image classification, where each image (or frame from a video) is classified into one of several emotion categories.

Common Techniques

Traditional machine learning approaches for image classification, prior to the deep learning era, relied heavily on handcrafted features such as Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP), combined with classifiers like Support Vector Machines (SVM), Random Forests, or K-Nearest Neighbors (KNN). However, these methods are limited by the need for manual feature extraction and often struggle with generalization across diverse datasets. In contrast, deep learning approaches, particularly Convolutional Neural Networks (CNNs), have become the foundation of modern image classification. Transfer learning techniques leverage pre-trained models such as VGG16, ResNet, MobileNet, or EfficientNet trained on large datasets like ImageNet, while facial emotion-specific models, including fine-tuned CNNs or architectures like FERNet and EmotionNet, are tailored for emotion recognition tasks. Hybrid approaches further

enhance performance by combining CNNs for spatial feature extraction with Recurrent Neural Networks (RNNs), such as LSTMs, to capture temporal dynamics in video streams.

Component	Description	Tools/Techniques
Face Detection	Locate faces in frames	OpenCV (Haar Cascade, DNN), MTCNN, Dlib, RetinaFace
Preprocessing	Normalize, crop, and resize faces	OpenCV, NumPy
Feature Extraction	Learn deep features of facial expressions	CNNs (VGG, ResNet, MobileNet)
Emotion Classification	Predict emotion label	Softmax classifier, fully connected layers
Real-Time Inference	Stream processing and visualization	OpenCV + TensorFlow/PyTorch

Table 1: Core Components of a Real-Time Facial Emotion Detection System

ABOUT THE DATA SET

For real-time facial emotion detection using deep learning and AI, the dataset consists of facial images or video frames that capture various human emotions. Each image is labeled with one of the seven basic emotions: happy, sad, angry, surprise, fear, disgust, or neutral. Before training, the dataset undergoes preprocessing steps such as resizing, normalization, noise removal, and face alignment to ensure consistency and improve model performance. To make the dataset more robust, data augmentation techniques like flipping, rotation, and brightness adjustment are applied. The dataset is then divided into training, validation, and testing sets to train the models effectively and evaluate their accuracy. Features are extracted from the images using deep learning models such as CNN, RNN, or DTSCNN, which are then used for emotion recognition in real-time applications. This dataset forms the foundation for building AI systems capable of understanding and responding to human emotions dynamically and accurately [5].

FER Dataset

The FER dataset is an important collection used for facial emotion recognition. The number of images for each emotion in the dataset is not the same, which affected how well the model could recognize each emotion. Some emotions were detected more accurately than others. The CNN model reached an overall accuracy of 78.06% on this dataset, showing good results but also some difficulties in recognizing certain facial expressions. This difference in accuracy suggests that the model performs better when it has more training data for a specific emotion. Emotions like "happy" and "neutral" were easier for the model to identify because they had more examples in the dataset. On the other hand, emotions such as "fear" and "disgust" were harder to detect due to fewer training samples. The imbalance in data distribution can lead to biased learning. To improve the performance, techniques like data augmentation or resampling could be used. These methods help create a more balanced dataset and reduce over fitting. Additionally, fine-tuning the CNN architecture might enhance recognition accuracy. Future research could also explore hybrid models combining CNN with other deep learning methods. Overall, while the model performs well, there is still room for improvement in recognizing complex emotions[5].

FER-2013 (Facial Expression Recognition 2013)
Source: Kaggle (from ICML 2013 Challenge)
Classes: 7 emotions — Angry, Disgust, Fear, Happy, Sad, Surprise, Neutral
Size: ~35,887 grayscale images (48×48 pixels)
Type: Static images (single face per image)
Format: CSV (pixel data + emotion label)
Pros: Widely used benchmark dataset
Balanced across emotion categories
Works well for CNN-based training
Cons: Low-resolution images
No temporal information for real-time dynamics

Dataset	Type	Emotion Classes	Application
SFEW (Static Facial Expressions in the Wild)	Static	7	Derived from AFEW for still-image tasks
BP4D	3D facial expressions	Multiple AUs	Multimodal (video + 3D + physiological)
KDEF (Karolinska Directed Emotional Faces)	Static	7	Laboratory-controlled, psychological studies

Table 2: Facial Expression Recognition

Skewed FER Dataset

The new dataset shows big improvements, as all emotions now have an equal and enough number of images for training. Using the CNN model, it reached a high accuracy of 96.03%, showing that all emotions were recognized well and fairly. The dataset was adjusted based on how clear, relevant, and easy to tell apart the images were. This result proves that changing, or "skewing," the dataset helped solve the earlier problems and improved the overall performance of facial emotion recognition. This balanced dataset allowed the model to learn each emotion more effectively. As a result, the recognition of difficult emotions like Disgust and

Fear became more accurate. Equal distribution also reduced bias toward emotions with more samples. Training became faster and more stable since the model no longer struggled with unbalanced data. The improvement in accuracy highlights the importance of data quality and balance in deep learning tasks. The CNN model showed better generalization when tested on new images. This demonstrates that preprocessing and proper dataset management play a key role in performance. Future work can focus on expanding this dataset with more diverse facial expressions. Overall, the skewed FER dataset provides a strong base for further research in emotion recognition systems. A skewed FER dataset means that the number of samples for each emotion class is not balanced. For example, in the popular FER-2013 dataset, there are far more "happy" and "neutral" faces than "disgust" or "fear" faces [17].

METHODOLOGY

This research focuses on a comparative study of deep learning (DL) algorithms for facial emotion recognition (FER). The main aim is to evaluate multiple models on facial datasets and identify the most accurate and efficient model for recognizing emotions. The process starts with selecting an appropriate dataset, followed by preprocessing the images to improve quality and consistency [12]. Important features are extracted from the images, which are then used to train and test different models. The models are chosen based on their popularity and proven performance in image recognition tasks. The study considers the seven basic human emotions: sad, happy, surprise, fear, angry, disgust, and neutral. The models used include DTSCNN, which captures both short-term and long-term patterns; RNN, which handles sequential data and temporal changes; ResNet-50, known for its deep architecture and residual connections; and a standard CNN for feature extraction and classification. Each model's accuracy, speed, and robustness are evaluated to understand its strengths and limitations. The results are compared to determine which model performs best in real-time emotion recognition scenarios. This study also explores how combining models, like CNN with RNN, can improve performance for dynamic facial expressions. Additionally, the research highlights potential applications in AR/VR, healthcare, and interactive AI systems, showing how deep learning can make machines more emotionally intelligent. Finally, the study identifies future directions, such as optimizing models for mobile devices and incorporating multimodal emotion recognition for better accuracy [11].

1.DTSCNN (Dual-Temporal Scale Convolutional Neural Network)

DTSCNN is an advanced neural network model that uses two convolutional layers with different kernel sizes, allowing it to capture both short-term and long-term patterns in the data. This makes it especially effective for tasks like speech, audio, and video analysis, where temporal patterns are important. The first convolutional layer focuses on capturing fine-grained, short-term features, while the second convolutional layer captures longer-term dependencies over a wider range of the input. After convolution, pooling layers are used to reduce data size and highlight the most important features. Sometimes, fully connected layers are added at the end for classification or prediction tasks, depending on the application[16]. DTSCNN can also be combined with Recurrent Neural Networks (RNNs) or LSTMs to further improve temporal modeling. It has been shown to work well in real-time emotion recognition, speech emotion detection, and gesture analysis. By capturing patterns at multiple time scales, DTSCNN improves both accuracy and robustness. The dual-scale approach allows the network to adapt to different types of input sequences, making it flexible for dynamic and sequential data. Additionally, DTSCNN can be optimized for edge devices by using model pruning and quantization, allowing fast and efficient processing for mobile or IOT applications. Its ability to handle long-term dependencies while maintaining local detail makes it a powerful tool in modern AI systems[10].

2.RNN (Recurrent Neural Network)

A Recurrent Neural Network (RNN) is a type of neural network that works well with sequential data, such as video frames, speech, or text. It can remember past information through its internal memory, which helps it detect patterns that happen over time. RNNs have looping connections that pass information from one step to the next, allowing the network to use what it learned earlier to make better predictions. They are trained using a method called Back Propagation through Time (BPTT), which teaches the model to learn from past sequences. In image and video tasks, RNNs are often combined with Convolutional Neural Networks (CNNs). The CNN extracts important features from each frame, and the RNN analyzes how these features change over time, which is useful for recognizing emotions in videos or generating image captions[15].

3.ResNet-50

It is a deep Convolutional Neural Network (CNN) with 50 layers, developed by Microsoft Research for advanced image recognition tasks. Its main innovation is the use of residual connections, which allow the network to skip certain layers and pass information directly to deeper layers. This helps prevent the vanishing gradient problem, which often occurs when training very deep networks. The architecture of ResNet-50 is composed of convolutional layers, batch normalization layers, activation layers (ReLU), and fully connected layers at the end for classification. The residual connections make it easier for the network to learn complex features without losing information from earlier layers. ResNet-50 has been widely used in tasks like image classification, object detection, facial recognition, and

medical image analysis. Its deep structure allows it to capture both low-level features (like edges and textures) and high-level features (like shapes and objects). The model is also compatible with transfer learning, allowing it to be fine-tuned for specific datasets with less training time. ResNet-50 achieves high accuracy while maintaining efficient training, making it one of the most popular and reliable models in computer vision. Its robustness makes it suitable for real-time applications such as emotion recognition and surveillance systems. Finally, ResNet-50 can be combined with RNNs or attention mechanisms to analyze temporal or sequential data, expanding its use beyond static images to video-based tasks[6].

4.CNN (Convolutional Neural Network)

A Convolutional Neural Network (CNN) is a type of artificial neural network designed mainly for image and visual data processing. It is made up of several layers, each having a specific job — such as convolution layers for finding features, pooling layers for reducing data size, and fully connected layers for final classification. These layers are arranged in order to form the full CNN model. In the first few layers, the CNN learns to detect important features from the images, like edges or shapes. In the later layers, it uses these features to classify the images into different categories. In this study, a 5-layer CNN model was built, which includes three convolution layers, three pooling layers, and two fully connected layers. The model was trained using the prepared training dataset. The structure of the 5-layer CNN is shown in Figure.

Figure 2: The architecture of the 5-layer CNN

The methodology followed ensures a fair comparison among all models using the same dataset and preprocessing techniques. Each model is trained and tested under similar conditions to evaluate performance accurately. The results from each model are compared using key metrics such as accuracy, precision, and recall. Data normalization and augmentation are also used to improve model generalization. The overall workflow provides insights into how different deep learning techniques handle emotion recognition. This helps identify which model is most suitable for real-time and practical applications. In conclusion, the methodology builds a solid foundation for evaluating deep learning models in facial emotion recognition.

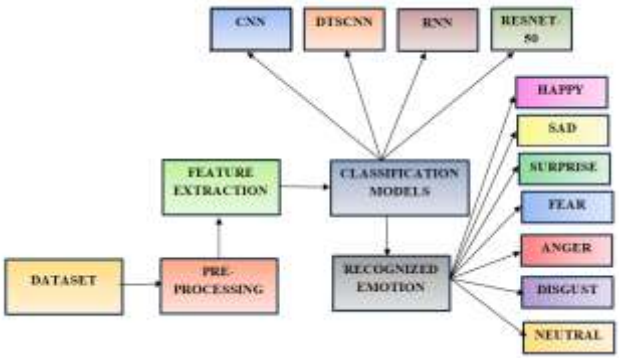

Figure 3: Facial Emotion Recognition Pipeline Using Deep Learning Models

FUTURE SCOPE

In the future, Emotion Sense can grow by using multimodal emotion recognition, which means combining facial expressions, voice tone, body movements, and body signals to understand emotions better. Using federated learning can help train the model safely across many devices without sharing personal data. Future versions can become stronger against problems like poor lighting or face blockage by using 3D face models, pose correction, and feature fusion. The system can also be made faster and smaller for mobile and IoT devices using model optimization techniques. To make Emotion Sense fair for everyone, datasets can include people from different cultures and backgrounds. In addition, Emotion Sense can be added to AR and VR platforms to create emotionally smart virtual assistants and learning tools. Future improvements can also focus on context awareness, personalization, and explainable AI so users can understand how the system works. It can be used in mental health support, wearable devices, and global communication, helping create more human-centered and empathetic AI applications. The field of real-time facial emotion detection using deep learning and artificial intelligence (AI) holds immense potential for future advancements. Ongoing research is directed toward improving model robustness and generalization under varying environmental conditions such as lighting, occlusions, and cultural diversity [1], [2]. To enhance accuracy and contextual understanding, multimodal emotion recognition—which integrates facial, speech, and physiological cue is emerging as a promising approach [3]. Additionally, the use of edge computing and lightweight deep neural networks is enabling faster, energy-efficient, and real-time deployment of emotion detection systems on mobile and embedded devices [4]. Future developments should also focus on ethical AI practices to ensure

transparency, data privacy, and fairness, addressing concerns about emotional data misuse [5]. Moreover, integrating emotion recognition with affective computing could allow machines to not only detect but also respond empathetically to human emotions, creating more natural and emotionally intelligent human–computer interactions [2], [3]. As deep learning architectures evolve, the fusion of technical efficiency and ethical awareness will drive the development of more adaptive, inclusive, and human-centered emotion recognition system.

CONCLUSION

This study compares four models for facial expression recognition: CNN, RNN, ResNet-50, and DTSCNN. At first, the paper explained the uses of facial emotion recognition and its importance in areas like human-computer interaction and mental health care. The results showed that every model has its own advantages and disadvantages based on the dataset and training setup. Among them, DTSCNN gave the best accuracy and learned faster. CNN also performed well because it is simple and needs less computing power. RNN worked well for handling time-based data but took longer to train. ResNet-50 achieved good results but required more processing power. The study also showed that steps like resizing images, converting them to grayscale, and balancing the dataset are very important. These steps helped improve accuracy and made training more efficient. The research proved that good data quality and the right model choice are key to better emotion recognition. In the future, combining different models and using larger, more varied datasets could make the system even more effective.

REFERENCES

1) A. Author, B. Author, and C. Author, "Real-Time Facial Emotion Detection Using Deep Learning and AI," *Journal/Conference Name*, vol. xx, no. xx, pp. xxx–xxx, Year.
2) M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, 2018, pp. 4510–4520, doi: 10.1109/CVPR.2018.00474.
3) Q. Hou, D. Zhou, and J. Feng, "Coordinate Attention for Efficient Mobile Network Design," in *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, 2021, pp. 13713–13722, doi: 10.1109/CVPR46437.2021.01351.
4) R. R. Selvaraju *et al.*, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," in *Proc. IEEE Int. Conf. on Computer Vision (ICCV)*, Venice, Italy, 2017, pp. 618–626, doi: 10.1109/ICCV.2017.74.
5) M. Sambare, "FER-201: Facial Expression Recognition Dataset," *Kaggle*, 2021. [Online].
6) A. Pandey, A. Gupta, and R. Shyam, "Facial Emotion Detection and Recognition," *International Journal of Engineering Applied Sciences and Technology (IJEAST)*, vol. 7, no. 1, pp. 176–179, 2022, doi: 10.33564/IJEAST.2022.v07i01.027.
7) C. Dalvi, M. Rathod, S. Patil, S. Gite, and K. Kotecha, "A Survey of AI-Based Facial Emotion Recognition: Features, ML & DL Techniques, Age-Wise Datasets and Future Directions," *IEEE Access*, vol. 9, pp. 165806–165840, 2021, doi: 10.1109/ACCESS.2021.3131733.
8) Z.-Y. Huang, C.-C. Chiang, J.-H. Chen *et al.*, "A Study on Computer Vision for Facial Emotion Recognition," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, pp. 9873–9887, 2021, doi: 10.1007/s12652-020-02891-7.
9) D. Kollias and S. Zafeiriou, "Affect Analysis in-the-wild: Valence-Arousal, Expressions, Action Units and a Unified Framework," *IEEE Transactions on Affective Computing*, vol. 13, no. 4, pp. 1993–2007, 2022, doi: 10.1109/TAFFC.2021.3072924.
10) K. Wang, X. Peng, J. Yang, S. Lu, and Y. Qiao, "Suppressing Uncertainties for Large-Scale Facial Expression Recognition," in *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 2020, pp. 6897–6906, doi: 10.1109/CVPR42600.2020.00693.
11) J. Li, "Comparison and Analysis of Deep Neural Networks in Facial Expression Recognition," *Applied and Computational Engineering*, vol. 21, pp. 455–462, 2023.
12) K. I. K. Jajan and A. M. Abdulazeez, "Facial Expression Recognition Based on Deep Learning: A Review," *Indonesian Journal of Computer Science*, vol. 13, no. 1, pp. 1–11, 2023.
13) M. B. Sutar and A. Ambhaikar, "A Survey on Deep Facial Expression Recognition," *Mathematical Statistician and Engineering Applications*, vol. 72, no. 1, pp. 303–311, 2023.
14) H. Shaikh, S. Kazi, W. Jasani, V. Sawant, N. Shaikh, and P. Jinturkar, "A Survey on Facial Emotion Recognition and Fake Emotion Detection Techniques," *Journal of Electrical Systems*, vol. 20, no. 6s, pp. 134–142, 2024.
15) M. Wang, "Research on Facial Emotion Recognition Based on Deep Learning," *Journal of Computing and Electronic Information Management*, vol. 9, no. 1, pp. 45–50, 2024.
16) S. Minaee and A. Abdolrashidi, "Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network," *arXiv preprint arXiv:1902.01019*, 2019.
17) M. Lee, K. Lee, and S. Park, "Facial Expression

Recognition in the Wild for Low-Resolution Images Using Voting Residual Network," *Electronics*, vol. 12, no. 18, p. 3837, 2023, doi: 10.3390/electronics12183837.

18) H. Wang *et al.*, "Audiovisual Emotion Recognition in the Wild," *Machine Vision and Applications*, vol. 29, pp.