

# **FER-Sys: A Real-Time Facial Emotion Recognition System Using Deepface And CNN Frameworks**

Akrisht Singh, Ravi Saharan

Department of Computer Science and Engineering, Central University of Rajasthan,  
Ajmer-305817-India.  
prsakrishtsingh@gmail.com  
ravisaharan@curaj.ac.in

## **Abstract**

Facial Emotion Recognition (FER) is a crucial aspect of developing emotionally aware human-computer interfaces. Existing models have achieved significant milestones but often lack flexibility or real-time performance. In this paper, we present FER-Sys, a comparative implementation of two approaches: the DeepFace framework and a custom-built Convolutional Neural Network (CNN) using Keras. These models were trained and evaluated on standard datasets, integrated into real-time webcam applications, and benchmarked for performance. The system's architecture enables facial alignment, emotion classification, and real-time prediction, with practical implications for emotion-aware applications.

**Keywords:** Facial Emotion Recognition, DeepFace, CNN, RealTime Detection, Human-Computer Interaction, Deep Learning.

## **1 Introduction**

Facial emotion recognition (FER) is an essential aspect of affective computing and computer vision, aiming to identify and interpret human emotions based on facial expressions. As facial cues are among the most direct and universally comprehensible forms of non-verbal communication, FER facilitates more intuitive interaction between humans and machines. It plays a critical role across various applications, including security systems, healthcare diagnostics, human-computer interaction, and behavioral analysis, by enabling systems to respond adaptively to human affective states[1].

The development of FER systems has evolved significantly over time. Early research was inspired by Charles Darwin's foundational work, *The Expression of the Emotions in Man and Animals* (1872), which proposed the universality of facial expressions across human cultures. Initial computational approaches employed handcrafted features such as Local Binary Patterns (LBP), Gabor filters, and facial landmarks to classify emotional states. These traditional methods often encountered limitations in dynamic, real-world environments due to variations in illumination, occlusions, and facial orientations. The advent of data-driven techniques, alongside the availability of annotated emotion datasets, enabled a shift towards more sophisticated and automated learning frameworks.

Recent advancements in deep learning have significantly enhanced the capabilities of FER systems. Convolutional Neural Networks (CNNs) have emerged as a standard architecture for image-based tasks due to their ability to extract spatial hierarchies of features efficiently. CNN-based models have demonstrated high accuracy in facial expression recognition by capturing subtle patterns and local variations across facial regions. DeepFace, developed by Facebook AI Research, is another influential deep learning framework initially designed for face verification and recognition. It has shown high generalization capabilities across facial identities and expressions, making it a viable candidate for FER tasks[2].

This paper is driven by the need to develop an efficient, accurate, and real-time FER system that can also address security concerns such as spoof detection. Many existing systems either lack real-time capabilities or perform inadequately under varying environmental conditions. The motivation behind this research lies in evaluating and comparing CNN and DeepFace-based models for FER, with a focus on their applicability

in robust CAPTCHA verification mechanisms. Through experimental analysis, the study aims to highlight the trade-offs and strengths of each approach, thereby contributing to the development of secure and intelligent emotion-aware systems.

### 1.1 Applications of FER

The FER systems have gained substantial attention due to their versatility and effectiveness across a wide range of realworld applications[3]. The following are some of the most prominent domains where FER plays a critical role:

- **Healthcare:** FER assists in diagnosing and monitoring mental health conditions such as depression, anxiety, and autism through emotional expression analysis.
- **Education:** FER evaluates student engagement and concentration levels in e-learning platforms to adapt instructional strategies accordingly.
- **Human-Computer Interaction:** FER enables emotionally aware interfaces in virtual assistants, chatbots, and interactive systems for improved user experience.
- **Reviews Marketing and Consumer Behavior:** FER analyzes emotional responses to products, services, or advertisements for data-driven branding and feedback collection.
- **Automotive Safety:** FER detects driver fatigue, distraction, or stress using in-cabin facial analysis to enhance road safety and prevent accidents.
- **CAPTCHA and Authentication Systems:** FER enhances traditional CAPTCHA systems by incorporating dynamic emotion-based facial challenges to differentiate between human users and bots.

## 2 Literature Review

Facial emotion recognition has been a longstanding area of interest within computer vision and human-computer interaction. Traditional methods relied on handcrafted features such as Local Binary Patterns (LBP) and Principal Component Analysis (PCA) to identify emotional states from facial expressions. However, these approaches struggled with generalization due to sensitivity to variations in lighting, pose, and occlusion [4][5].

The emergence of deep learning revolutionized FER systems. Convolutional Neural Networks (CNNs) proved effective in learning hierarchical and spatial features from facial images, significantly improving classification accuracy. Mellouk and Handouzi provided a comprehensive overview of deep learning techniques for FER, highlighting the superiority of CNNs over traditional approaches in terms of accuracy and robustness [6]. Kanagaraju et al. also demonstrated the effectiveness of image processing-based deep learning models in capturing fine-grained emotional features under diverse conditions [7].

DeepFace, introduced by Facebook AI Research, was originally developed for face verification but has been adapted for emotion recognition due to its robust feature extraction capabilities and use of 3D face alignment. Zhang evaluated DeepFace's effectiveness in FER tasks, showing its potential for high generalization across identities and expressions [8]. Other methods like wavelet energy features [9], local binary patterns [4], and ensemble models such as boosting algorithms [10] have also contributed to FER advancements.

Mukhopadhyay et al. explored the application of FER in online learning environments, emphasizing the need for real-time and accurate recognition to assess students' affective states [11]. Despite these advancements, limited research has focused on comparative analysis between traditional CNNs [13] and advanced pre-trained models like DeepFace, especially in practical, real-time applications [15] such as CAPTCHA-based authentication systems. This paper aims to fill that gap by experimentally comparing both approaches under a common evaluation framework.

## 3 Methodology

This study implements and compares two facial emotion recognition approaches: a custom Convolutional Neural Network (CNN) and a pre-trained DeepFace model. Both systems are trained and tested on the FER2013 dataset and evaluated for accuracy, real-time performance, and suitability for integration into user-

facing applications.

### 3.1 Dataset Used

The proposed EECVS leverages DeepFace for emotion recognition, which employs a VGG-Face-based model pre-trained on the Facial Expression Recognition (FER2013) dataset [10]. This dataset contains a diverse set of facial expressions crucial for training robust emotion classifiers. Fig. 5 illustrates the corresponding emoji representations used in the system to prompt user expressions during verification.

Validation is performed with dataset comprising of 7 different types of emotion images. Table-II provide details of each emotion tag instance counts used for proposed system validation. Real-time testing is done using a live camera to evaluate system performance and anti-spoofing[11].

**Table 1.** Instance count of Emotion Dataset

Emotion Tag	Instance Count
Happy	7164
Sad	3205
Angry	4938
Surprise	4982
Neutral	4103
Disgust	436
Fear	3993

### 3.2 CNN-Based Facial Emotion Recognition

A custom Convolutional Neural Network was developed using the Keras library with TensorFlow backend. The architecture consists of:

- Three convolutional blocks, each followed by batch normalization, ReLU activation, and max pooling.
- Two fully connected layers with dropout for regularization.
- A final softmax output layer for classifying into one of the seven emotion categories.

The model is trained using the Adam optimizer with categorical cross-entropy loss. Early stopping and model checkpoint callbacks are employed to prevent overfitting and retain the best performing model.

### 3.3 DeepFace-Based Facial Emotion Recognition

DeepFace, a pre-trained face recognition model developed by Facebook AI, is adapted for emotion classification using the VGG-Face backbone. The model performs the following operations:

- Face detection and alignment using OpenCV and dlib for preprocessing.
- Feature extraction using the DeepFace embedding layer.
- Classification into one of the seven emotion classes based on a fine-tuned dense head.

The model benefits from transfer learning and requires minimal retraining, providing high accuracy with lower computational cost.

## 4 Architecture

#### 4.1 Preprocessing

711

- **Face Detection:** OpenCV is used to detect and localize the face in the image.
- **Face Alignment:** The detected face is aligned to ensure consistent orientation of facial features.
- **Image Enhancement:** Techniques like histogram equalization are applied to improve image quality.

#### 4.2 DeepFace-Based Facial Emotion Recognition Model

- **Feature Extraction:** Pre-trained DeepFace model VGGFace is used to extract facial embeddings representing key features.
- **Emotion Classification:** The embeddings are passed through a classifier to predict emotions (e.g., happy, sad, angry) with associated confidence scores.

#### 4.3 DeepFace-Based Facial Emotion Recognition Model

- **Feature Extraction:** Convolutional layers extract facial features from the image.
- **Emotion Classification:** Fully connected layers output emotion predictions (e.g., happy, sad, angry) with confidence scores.

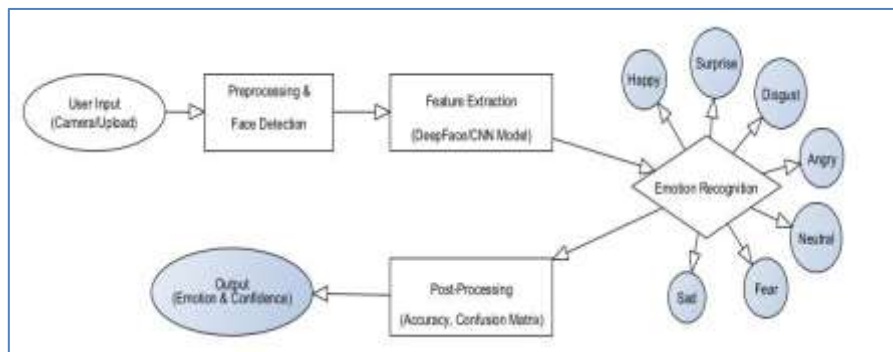


Fig. 1. Flow Diagram of FER-Sys Real-Time Prediction

## 5 Experimental Setup

To evaluate the performance of the DeepFace and CNN based emotion recognition models, a controlled experimental environment was established.

### 5.1 Dataset

Experiments were conducted using the FER-2013 dataset, a widely adopted benchmark for facial emotion recognition. The dataset comprises 28,821 grayscale facial images of size 48×48 pixels, each labeled with one of seven emotion classes: Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral. The dataset was divided into training (70%), validation (15%) and testing(15%) subsets. Data augmentation techniques, such as rotation, zooming, and horizontal flipping, were applied to enhance model generalization.

### 5.2 Model System Configuration:

- **Hardware:** Intel i7 CPU, 16GB RAM, NVIDIA GTX 1660 Ti
- **Operating System:** Windows 11
- **Programming Environment:** Python 3.8, with supporting libraries including TensorFlow, Keras, OpenCV, and DeepFace.

### 5.3 Model Implementation

- **DeepFace Model:** The pre-trained DeepFace framework was employed, leveraging the VGG-Face backbone for feature extraction and emotion classification. The model was used in an inference-only mode without additional training.
- **CNN Model:** A custom Convolutional Neural Network was developed and trained on the above dataset with epoch of CNN model to be 30. The architecture includes three convolutional, pooling, and two fully connected layers, concluding with a softmax classifier. The model was optimized using the Adam optimizer and categorical cross-entropy loss.

## 6 Results and Analysis

### 6.1 Performance Metrics

The DeepFace model significantly outperformed the CNN model across all key metrics. Notably, DeepFace achieved over 82% in both macro and weighted F1-scores, indicating robust and consistent classification across all emotion categories. In contrast, the CNN model struggled, particularly with minority classes like Disgust, which yielded an F1-score of just 0.04

**Table 2.** Performance comparison for CNN and DeepFace Model

Metric	CNN	DeepFace
Accuracy	0.59	0.82
Macro Average F1	0.48	0.82
Weighted Average F1	0.57	0.82
Best Class(F1)	Happy(0.83)	Fear(0.89)
Worst Class(F1)	Disgust(0.04)	Angry(0.76)

### 6.2 Real-time Prediction:

We have Developed an interactive interface as shown in Fig.2, where users can upload images for emotion prediction using the DeepFace model.



**Fig. 2.** Sample image of Real-time Prediction

### 6.3 Visualization and Evaluation

To complement the quantitative evaluation, several visual tools were employed to better understand and compare the behavior of the CNN and DeepFace models across different emotion classes.

#### 6.3.1 Accuracy Comparison Graph

Fig.3 illustrates a comparative line chart of the classification accuracy for each of the seven emotion categories (Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral). The DeepFace model demonstrated consistently higher accuracy across all categories. Surprise (F1-score: 91.42%) and Fear (F1-score: 89.31%) performed the best, showing high detection consistency. Angry (76.66%) and Disgust (76.58%) were less accurate, likely due to subtle expression overlap. Notably, Fear had the highest precision (91.74%), while Neutral showed better recall (82.86%) than precision. The CNN model showed competitive performance only in the Happy category(83%), where it achieved its highest class-specific accuracy.

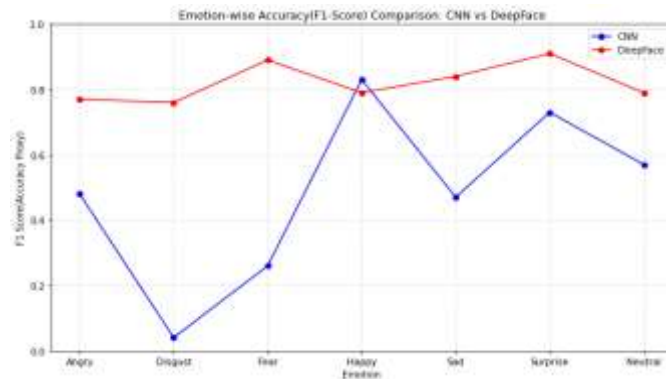


Fig. 3. Accuracy Comparison Graph

### 6.3.2 Precision and Recall Comparison

A side-by-side bar chart (Fig.4) compares the precision and recall of DeepFace and CNN models across all emotion classes. The DeepFace model consistently achieved higher values in both metrics, reinforcing its reliability and robustness. The CNN model displayed greater variance, with lower scores particularly in minority classes such as Disgust and Fear.

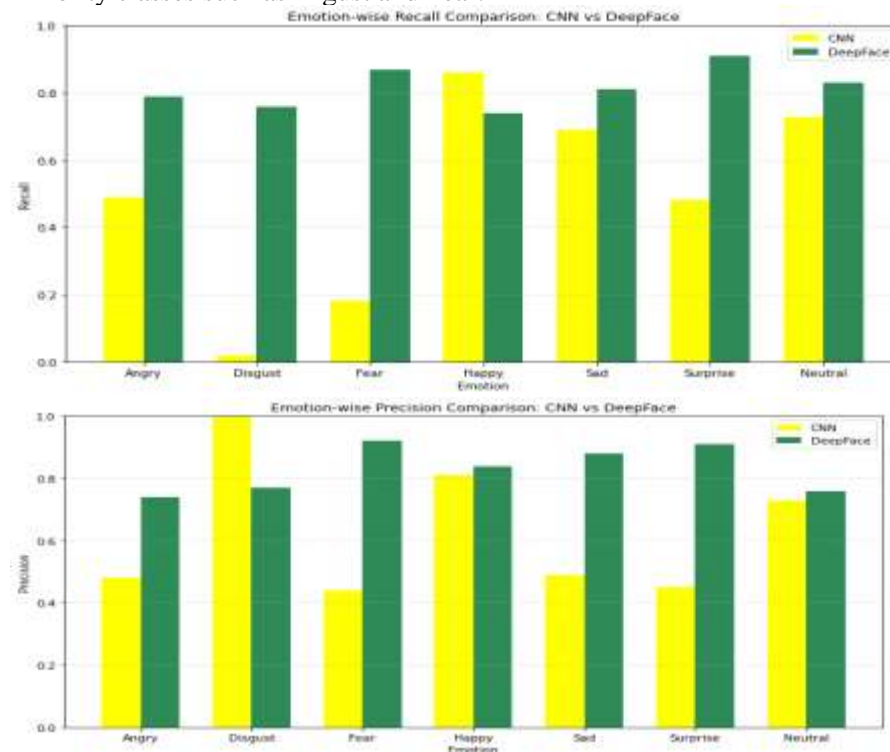


Fig. 4. Precision and Recall bar chart

### 6.3.3 Confusion Matrix

Fig. 5 illustrates the confusion matrix for the DeepFace model, showing strong diagonal dominance and high classification accuracy, especially for Surprise (91.42%) and Fear (89.31%). Misclassifications are minimal, with slight confusion between visually similar classes such as Angry and Disgust. The results demonstrate balanced and consistent performance, confirming DeepFace’s effectiveness for real-time emotion-aware systems like CAPTCHA verification and interactive agents.

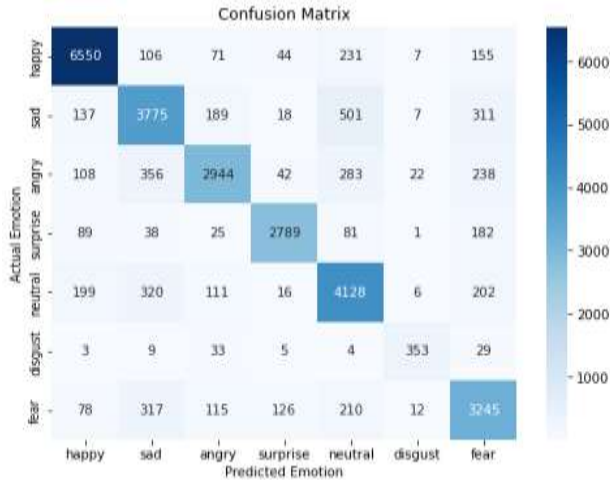


Fig. 5. DeepFace model Confusion Matrix

Fig. 6 presents the confusion matrix for the custom CNN model, highlighting its limitations in handling minority and subtle emotion classes. While it performs reasonably well in classifying Happy, it shows significant misclassifications for emotions like Disgust, as reflected in the sparse diagonal and distributed off-diagonal entries. These shortcomings are likely due to class imbalance, limited network depth, and insufficient feature alignment.

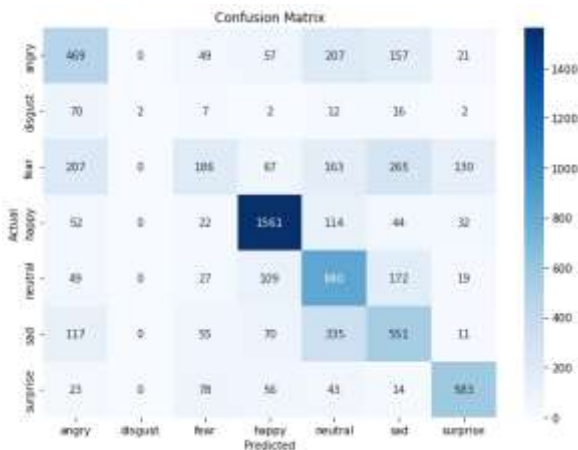


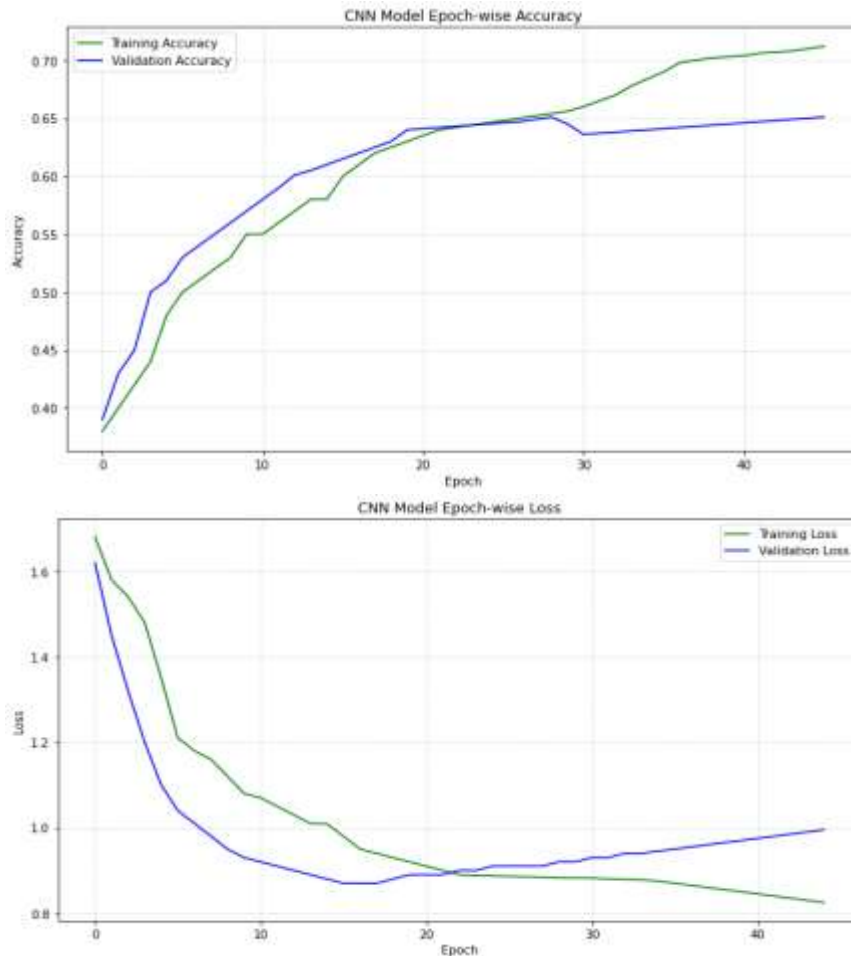
Fig. 6. CNN based model Confusion Matrix

### 6.3.4 Accuracy and Loss Metrics

Fig.7 visualizes the CNN training process over multiple epochs. The accuracy curve shows an initial rise followed by early stabilization, suggesting that the model reaches a plateau without further significant improvement. Meanwhile, the loss curve gradually declines, indicating learning but possibly underfitting due to architecture constraints or insufficient learning capacity. This trend implies that deeper networks or regularization enhancements (e.g., dropout tuning or learning rate adjustments) may be needed for better

generalization.

715



**Fig. 7.** CNN Model Accuracy and Loss Metrics

## 7 Conclusion and Future Scope

In this paper, we implemented and evaluated two facial emotion recognition models: a custom Convolutional Neural Network (CNN) and a pre-trained DeepFace model based on VGG-Face. Experimental results on the FER-2013 dataset revealed that the DeepFace model significantly outperformed the CNN across all key performance metrics, including accuracy, precision, recall, and F1-score. DeepFace demonstrated strong generalization and consistent performance across all emotion classes, particularly excelling in complex expressions such as Fear(89%) and Sad(84%), while the CNN model exhibited comparatively lower accuracy and struggled with minority classes like Disgust. Despite its inferior performance, the CNN model offers advantages in flexibility and can be further optimized. In future work, we aim to enhance the CNN architecture using transfer learning and deeper layers, explore real-time deployment strategies, and incorporate multimodal data for improved emotion detection. Additionally, expanding the training dataset and optimizing for edge devices will be prioritized to increase model robustness and practical applicability in user-centric environments such as e-learning, healthcare, and adaptive interfaces.

## References

1. J. Kaur, J. Saxena, J. Shah, Fahad and S. P. Yadav, "Facial Emotion Recognition," 2022 International Conference

- on Computational Intelligence and Sustainable Engineering Solutions (CISES), Greater Noida, India, 2022, pp. 528-533, doi: 10.1109/CISES54857.2022.9844366.
2. R. Saharan and N. Kesswani, "A Method to Recognize Facial Emotion for Crowded Images," 2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI), Coimbatore, India, 2024, pp. 858-863, doi: 10.1109/ICoICI62503.2024.10696288.
  3. Chowdary, M.K., Nguyen, T.N. Hemanth, D.J. Deep learning-based facial emotion recognition for human-computer interaction applications. *Neural Comput Applic* 35, 23311–23328 (2023).
  4. Zhao, L.; Zhuang, G.; Xu, X. Facial expression recognition based on PCA and NMF. In Proceedings of the 2008 7th World Congress on Intelligent Control and Automation, Chongqing, China, 25–27 June 2008; pp. 6826–6829.
  5. Xiao, X.Q.; Wei, J. Application of wavelet energy feature in facial expression recognition. In Proceedings of the 2007 International Workshop on Anti Counterfeiting, Security and Identification (ASID), Xiamen, China, 16–18 April 2007; pp. 169–174.
  6. Mellouk, W., and Handouzi, W. "Facial emotion recognition using deep learning: review and insights," *Procedia Computer Science*, 2020.
  7. Zhang, Z. "Deep Face Emotion Recognition," *Journal of Physics: Conference Series*, 2018.
  8. Pandey, Amit Gupta, Aman Shyam, Radhey. (2022). FACIAL EMOTION DETECTION AND RECOGNITION. 10.33564/IJEAST.2022.v07i01.027.
  9. Kanagaraju, P. et al. "Emotion detection from facial expression using image processing," *International Journal of Health Sciences*, 2022.
  10. Mukhopadhyay, M. et al. "Facial Emotion Detection in Online Learning Systems," *ACM*, 2020.
  11. Feng, X.; Pietikainen, M.; Hadid, A. Facial expression recognition based on local binary patterns. *Pattern Recognit. Image Anal.* 2007, 17, 592–598.
  12. Zhi, R.; Ruan, Q. Facial expression recognition based on twodimensional discriminant locality preserving projections. *Neurocomputing* 2008, 71, 1730 1734.
  13. Lee, C.-C.; Shih, C.-Y.; Lai, W.-P.; Lin, P.-C. An improved boosting algorithm and its application to facial emotion recognition. *J. Ambient. Intell. Humaniz. Comput.* 2012, 3, 11–17.
  14. FER2013 Dataset, Kaggle. Available: <https://www.kaggle.com/datasets/deadskull7/fer2013>
  15. Chang, C.-Y.; Huang, Y.-C. Personalized facial expression recognition in indoor environments. In Proceedings of the 2010 International Joint Conference on Neural Networks (IJCNN), Barcelona, Spain, 18–23 July 2010; pp. 1–8.