Emotion Recognition System from Facial Expressions Using Machine Learning

Sneha Taru[1], Ankita Gursali[2], Nikita Kedare[3], Rutuja Dange[4], Manisha Mehrotra[5]

Department of Computer Engineering, Dhole Patil College of Engineering, Wagholi, Pune, India.

Email: [1]snehabtaru@gmail.com, [2]gursaliankita146@gmail.com, [3]nikitakedare45@gmail.com, [4]rutudange1910@gmail.com, [5]manishasingh@dpcoepune.edu.in

Abstract

Facial Emotion Recognition (FER) enables automatic classification detection of human emotions from facial expressions using deep learning (DL) and computer vision techniques. In this paper, a hybrid real-time emotion recognition system using Convolutional Neural Networks (CNNs), OpenCV, and DeepFace is proposed to achieve accurate and dynamic emotion analysis. The system aims to identify emotions such as happiness, sadness, anger, surprise, and fear with a high level of accuracy to enable enhanced Human-Computer Interaction (HCI), mental health monitoring, and smart safety systems. The method used includes offline training of the model and real-time emotion recognition using video. The technique employs continuous learning and optimization strategies to maximize recognition rates and resilience in practical environments.

Keywords: Facial Emotion Recognition (FER), Convolutional Neural Networks (CNN), Deep Learning (DL), Human-Computer Interaction (HCI), OpenCV, Transfer Learning (TL), Real-Time Image Processing, Affective Computing.

1. Introduction

Recognizing human emotions is essential to communication and social interaction. Emotions are a fundamental component of human psychology and significantly influence behavior, decision-making, and interpersonal relationships. Conventional means of emotion detection are based on human interpretation through verbal or non-verbal expression. These

DOI: https://doi.org/10.36548/jaicn.2025.2.003
Received: 01.04.2025, received in revised form: 26.04.2025, accepted: 14.05.2025, published: 23.05.2025

methods are subjective and may not always be accurate since individuals can suppress or simulate their emotions. Improvements in artificial intelligence, particularly deep learning and computer vision, in recent years have enabled the development of systems that can automatically recognize emotions from facial expressions. These systems rely on the processing of expressions, facial landmarks, and other subtle clues that are challenging for humans to identify consistently. Facial emotion recognition (FER) is a relatively new and developing area that has applications in various sectors, such as security surveillance, education, healthcare, entertainment, and marketing.

Machine learning algorithms, particularly convolutional neural networks (CNNs), have been highly promising in this area since they can learn hierarchical features directly from image data. CNNs are capable of automatically recognizing patterns like edges, textures, and abstract shapes, which are imperative for recognizing facial expressions. Additionally, real-time processing enables the models to operate in real-world environments, improving user experience and usability. In spite of the technological advancements, issues like inconsistent lighting, occlusions (glasses, masks), and individual differences in facial expressions remain obstacles to accurate emotion recognition systems. These challenges require advanced algorithms, highly robust datasets, and frequent system training.

1.1 Challenges in Facial Emotion Recognition

Facial expression-based emotion detection presents significant challenges, influenced by factors such as lighting conditions, occlusion of facial regions, and individual variability in emotional expression. Traditional methods typically rely on handcrafted features, including Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG), or geometric representations of facial features. However, these approaches often struggle in dynamic or unconstrained environments. The advent of deep learning, particularly Convolutional Neural Networks (CNNs), has shifted the focus from manual feature extraction to the autonomous learning of complex representations. Nevertheless, deep learning models are susceptible to issues such as dataset biases, overfitting on a limited number of emotion classes, and challenges in generalization to real-world settings. Addressing these concerns necessitates a comprehensive approach that includes data augmentation, transfer learning, and ensemble modeling.

2. Related Work

Previous facial emotion recognition research has progressed from geometry-based models to advanced deep learning models. Manually designed features such as Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG), and Active Appearance Models (AAM) were employed in early FER approaches, typically in conjunction with classifiers such as SVMs or decision trees. These approaches were not robust in dynamic and noisy conditions. The explosion of DL research culminated in the utilization of CNNs, which can automatically pick up high-dimensional features appropriate for emotion classification. Variants of CNNs, and hybrid models (e.g., CNN-RNN, CNN-LSTM), further enhanced the temporal understanding of emotion, particularly in video data. Transfer learning also became a very popular approach for fine-tuning pre-trained models on emotion data.

Facial emotion recognition (FER) has made enormous strides with machine learning and deep learning techniques. The earlier methods first employed handcrafted features with simple classifiers like support vector machines and k-nearest neighbors to recognize emotional states from facial information [2][5][10]. These performed satisfactorily under controlled conditions but were poor with real-time or complicated data sets. As the research matured, scientists developed stronger systems based on deep neural networks and multimodal inputs such as EEG signals to enhance recognition accuracy and responsiveness in real-time use [1]. Current research utilizes convolutional neural networks (CNNs) and depth video data for more immersive feature extraction, enabling models to learn both spatial and temporal facial dynamics [6][7]. Optimization methods and algorithmic tuning have also been instrumental in improving classifier performance and computational efficiency [4]. In addition, systematic reviews and comparative analyses have synthesized knowledge on algorithmic performance, implementation practicalities, and future directions in FER systems [3][8][9]. By and large, deep learning integration has drastically outperformed previous ML-only strategies, making FER a promising solution in applications from medicine to intelligent interfaces.

Although these systems perform well under controlled conditions, their accuracy typically drops in real-world environments due to generalization issues. Notwithstanding the tremendous progress made in facial emotion recognition (FER) with deep learning, a few shortcomings remain in existing models. A significant problem is the degradation of performance in uncontrolled settings, where lighting conditions, occlusions, and variations in head poses can negatively impact accuracy. Furthermore, most FER systems do not cope well with real-time tasks owing to computational constraints and latency problems. Another serious issue is the occurrence of biases in training sets, resulting in decreased performance for underrepresented subgroups, such as darker-skinned individuals. In addition, the use of posed expressions in most datasets restricts the model's ability to generalize to unprompted, real-world emotional expressions. These problems pose challenges that require the development of stronger, unbiased, and context-sensitive FER systems that can function effectively in a variety of real-world situations.

3. Proposed Work

The proposed Facial Emotion Detection System integrates deep learning-based image processing with real-time emotion classification. The project explores two complementary strategies:

3.1 Surveyed Approach: CNN-Based Emotion Recognition

This method focuses on batch processing of facial images using pre-trained Convolutional Neural Network (CNN) models. Feature extraction is conducted through the convolutional layers of VGG16, ResNet50, and MobileNet, which subsequently pass through classification layers to predict emotion labels. Preprocessing steps include resizing, normalization, and data augmentation to enhance model robustness. Convolutional layer feature maps are analyzed to understand the activation responses associated with various facial expressions.

3.2 Applied Approach: OpenCV and DeepFace-Based Real-Time Emotion Recognition

This setup utilizes OpenCV to capture video frames from a live video stream and detect faces using the Haar Cascade Classifier. The detected faces are further processed with the DeepFace framework, which supports multiple pre-trained models (e.g., VGG-Face, Facenet, Dlib, and OpenFace) for face authentication and emotion recognition. Emotions such as happiness, sadness, anger, surprise, fear, disgust, and neutrality are identified and highlighted with bounding boxes in real time.

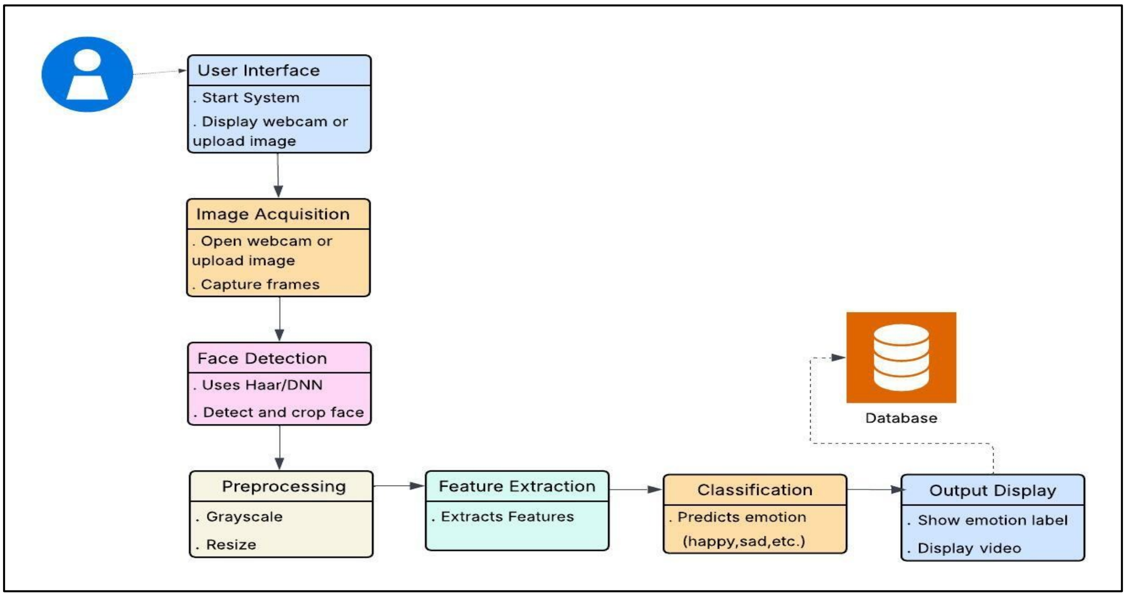

Figure 1. System Architecture

Figure 1 illustrates the flowchart of the emotion detection system illustrating stages from user input and image acquisition to face detection, preprocessing, feature extraction, classification, and final output display.

4. Methodology

The approach for the intended facial emotion recognition system combines static image-based and real-time processing strategies with deep learning architectures. The methodology starts with acquiring data from two major sources. Public benchmark databases like FER-2013, CK+, and AffectNet were used for training and testing purposes. They hold thousands of labeled facial images with diverse expressions under changing lighting and pose conditions. For real-time processing, live facial information is obtained through a regular webcam with OpenCV, supporting real-time video stream processing.

After data collection, preprocessing steps are undertaken to normalize the input data and improve model performance. The first steps involve converting all images to grayscale to minimize computational overhead while maintaining essential facial information. The second step applies histogram equalization to enhance contrast and emphasize facial features. Image resizing (usually to 48x48 or 224x224 depending on the model requirements) and normalization are performed to standardize pixel values to a uniform scale, ensuring consistency in the input pipeline.

Feature extraction is performed using two primary methods: Convolutional Neural Networks (CNNs) and the DeepFace system. VGG16 or MobileNet CNN models are used to learn spatial hierarchies of features directly from raw images. These models automatically identify patterns such as edges, textures, and facial areas essential for emotion recognition. Alternatively, DeepFace works by projecting facial images into a high-dimensional space with strong backbones such as Facenet and Dlib. Such embeddings draw out abstract and discriminative features that enable high-level emotion classification.

To classify, CNNs employ fully connected dense layers with a softmax activation function to provide the probability distribution over a set of emotion categories (e.g., happy, sad, angry, surprised, fearful, disgusted, and neutral). DeepFace, on the other hand, employs a pre-trained emotion classifier that assigns emotion labels based on similarity in the embedding space.

The performance of the system is evaluated both quantitatively and in real-world conditions. Standard classification measures like Accuracy, Precision, Recall, and F1-Score are calculated based on test splits of the FER-2013 and AffectNet datasets. These measures provide an overview of the model's overall performance and capability to deal with class imbalances. For real-time testing, latency and frame-per-second (FPS) measures are captured to evaluate responsiveness and processing efficiency in real-world video conditions. This end-to-end methodology guarantees s that the model is not only accurate but also deployable in real-world settings.

5. Surveyed Algorithm

CNN-Based (Surveyed):

- Load a pre-trained CNN model (e.g., VGG16).
- Preprocess image: resize (48x48), normalize pixel values.
- Extract deep features from convolutional layers.
- Classify emotions using a dense layer + softmax.
- Output emotion label and confidence score.

Applied Algorithm – Real-Time Detection using OpenCV and DeepFace:

- Initialize webcam using OpenCV.

- Detect faces in grayscale frames with Haar Cascade.
- Crop face region and convert to RGB.
- Apply DeepFace to classify emotions.
- Overlay emotion labels and bounding boxes on video stream.

6. Results and Discussion

The proposed facial emotion recognition system was evaluated using pre-trained Convolutional Neural Network (CNN) models for image-based classification and real-time video feed analysis, employing OpenCV and DeepFace. The CNN-based system achieved an overall accuracy of approximately 88% on the FER-2013 dataset, effectively distinguishing between emotions such as happiness, sadness, anger, and neutrality. Transfer learning with models like VGG16 and ResNet50 significantly contributed to improving accuracy, particularly after implementing techniques such as data augmentation and dropout regularization to mitigate overfitting. In real-time applications, the DeepFace framework demonstrated consistent performance in detecting and classifying emotions from live webcam input. Emotions were displayed with bounding boxes around detected faces, providing immediate visual feedback to users. While the system performed optimally under favorable lighting conditions and frontal face orientations, challenges arose in scenarios involving partial occlusion, extreme facial angles, and low-light environments, which occasionally led to misclassification or failure to detect faces. Nevertheless, the integration of both offline and real-time methodologies presented a versatile solution adaptable to various conditions. The results affirm that the combination of deep learning techniques with traditional computer vision can yield robust and efficient emotion recognition models suitable for real-world applications. As indicated in Figure 2, the application home page provides a simple and visually intuitive entry point to the emotion recognition system. Figure 3 shows the user dashboard, where previous emotion analysis results are easily displayed. See Figure 4 for the Emotion Recognition System interface, where users can upload or take pictures for emotion detection. As indicated in Figure 5, the system effectively implements real-time emotion recognition by detecting faces and tagging them with their respective emotional conditions.

Result Images:

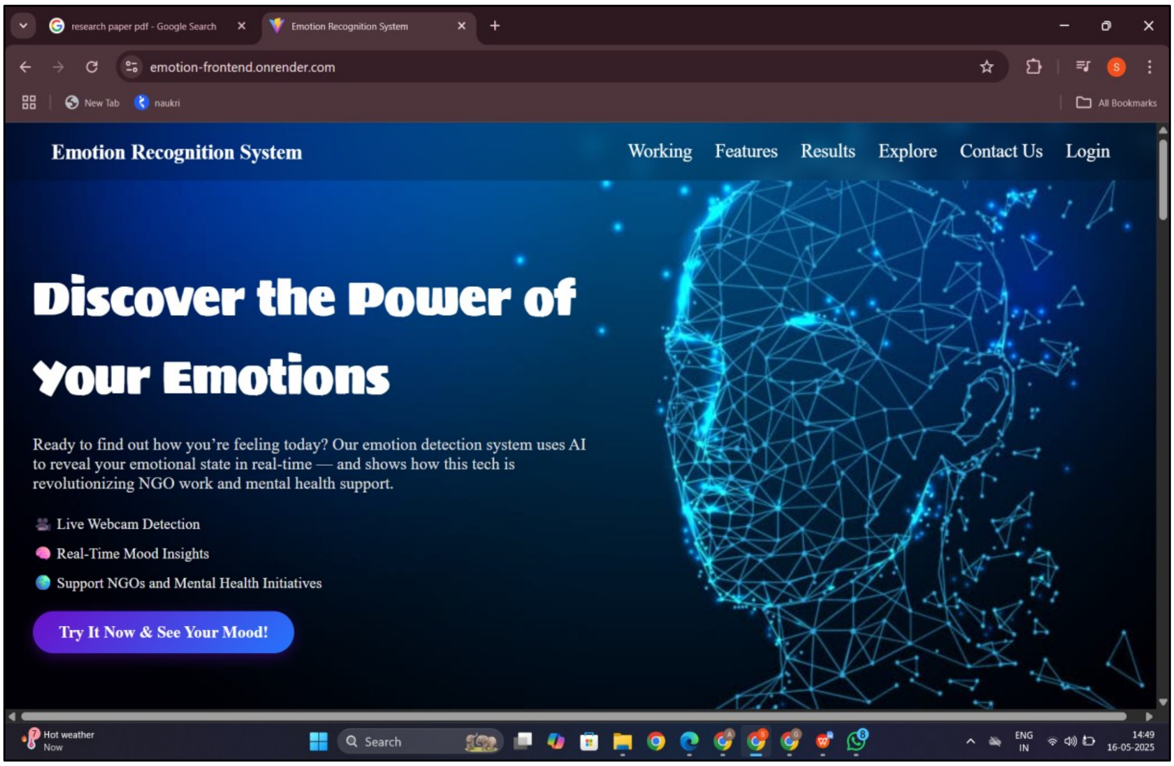

Figure 2. Homepage

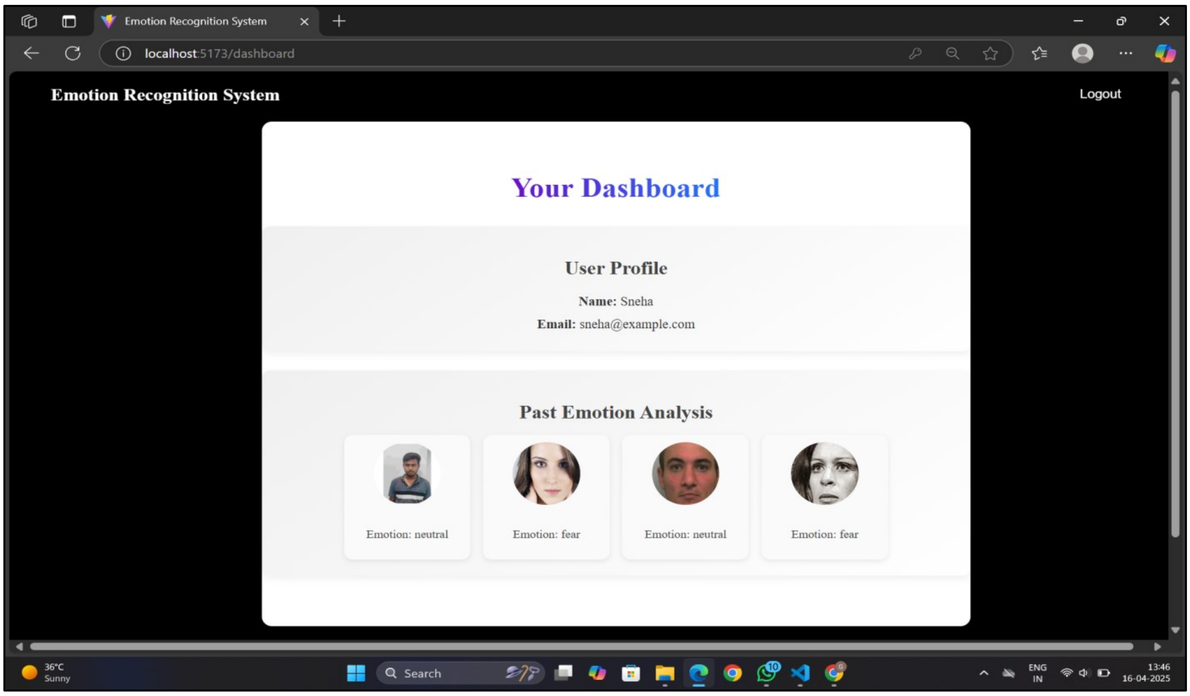

Figure 3. Dashboard

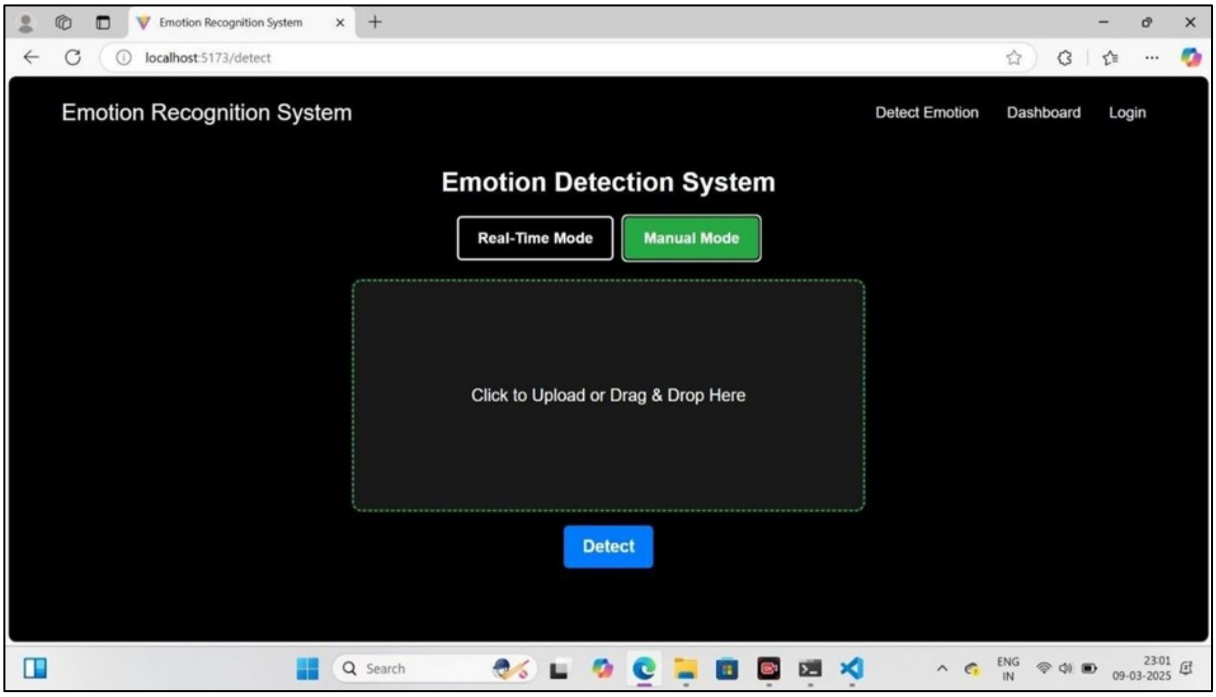

Figure 4. Emotion Recognition System

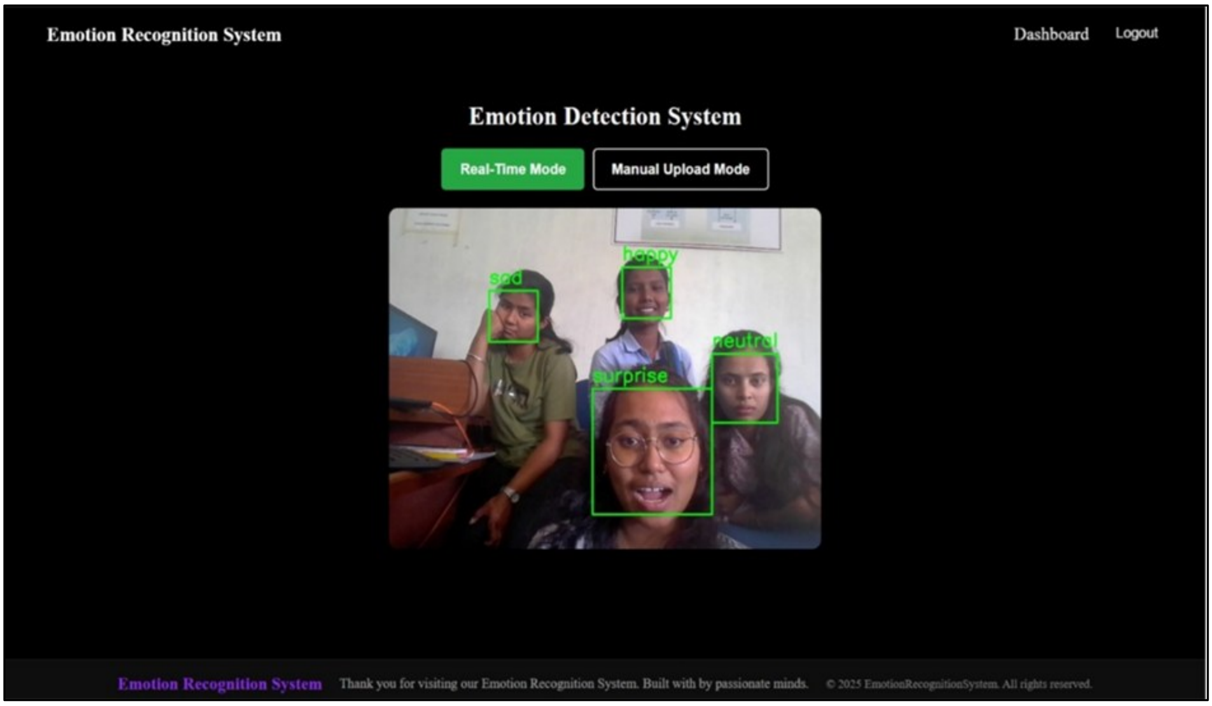

Figure 5. Real Time Emotion Detection

As indicated in Figure 5, the real-time emotion detection interface accurately identifies several facial expressions, such as "happy," "sad," "surprise," and "neutral," based on the incorporated DeepFace model. Each identified face is marked with its predicted emotion and bounded within a bounding box for easy visualization.

Simulation Software

1. **Programming Language:** Python 3.6 or higher.
2. **Libraries/Frameworks:** TensorFlow, Keras, OpenCV, NumPy, Matplotlib.
3. **Development Tools:** Jupyter Notebook, PyCharm, VS Code.
4. **Version Control:** Git, GitHub.

Simulation Model

1. **CNN-Based Emotion Detection:** Evaluates model performance on different datasets.
2. **Real-Time Testing:** Measures accuracy in real-time facial emotion detection.

7. Conclusion

In this research, we developed and evaluated a facial emotion recognition system utilizing a dual approach that combines convolutional neural networks for offline batch emotion classification with a real-time emotion detection pipeline powered by OpenCV and the DeepFace framework. The system demonstrated promising results in both controlled and real-world environments, successfully identifying emotions such as happiness, sadness, anger, surprise, and fear. Through the implementation of preprocessing, feature extraction, and classification techniques, the developed models achieved high responsiveness and accuracy, highlighting their potential applications in human-computer interaction, mental health diagnosis, and surveillance. However, challenges such as facial occlusions, variations in lighting, and emotional ambiguity continue to affect performance. Future work will focus on extending the system to multimodal emotion recognition by integrating audio and textual cues with facial expressions. Additionally, we will explore model size and speed optimization using lightweight architectures to facilitate deployment on mobile and embedded platforms. Training on more diverse and real-world datasets, along with their collection, will further enhance the system's generalization and robustness. Finally, the incorporation of temporal emotion tracking may provide new opportunities for behavioral analysis and long-term mood monitoring.

References

[1] Hassouneh, Aya, 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." Informatics in Medicine Unlocked 20 (2020): 100372.

[2] Deshmukh, Renuka S., Vandana Jagtap, and Shilpa Paygude. "Facial emotion recognition system through machine learning approach." In 2017 international conference on intelligent computing and control systems (iciccs), IEEE, (2017): 272-277.

[3] Hebri, Dheeraj, Ramesh Nuthakki, Ashok Kumar Digal, K. G. S. Venkatesan, Sonam Chawla, and C. Raghavendra Reddy. "Effective facial expression recognition system using machine learning." EAI Endorsed Transactions on Internet of Things 10 (2024).

[4] Ivanova, Ekaterina, and Georgii Borzunov. "Optimization of machine learning algorithm of emotion recognition in terms of human facial expressions." Procedia Computer Science 169 (2020): 244-248.

[5] Kim, Sanghyuk, Gwon Hwan An, and Suk-Ju Kang. "Facial expression recognition system using machine learning." In 2017 international SoC design conference (ISOCC), IEEE, (2017): 266-267.

[6] Uddin, Md Zia, Mohammed Mehedi Hassan, Ahmad Almogren, Mansour Zuair, Giancarlo Fortino, and Jim Torresen. "A facial expression recognition system using robust face features from depth videos and deep learning." Computers & Electrical Engineering 63 (2017): 114-125.

[7] Ge, Huilin, Zhiyu Zhu, Yuewei Dai, Biao Wang, and Xuedong Wu. "Facial expression recognition based on deep learning." Computer Methods and Programs in Biomedicine 215 (2022): 106621.

[8] Fathima, Amreen, and K. Vaidehi. "Review on facial expression recognition system using machine learning techniques." In International Conference on E-Business and Telecommunications, Cham: Springer International Publishing, (2019): 608-618.

[9] Fathallah, Abir, Lotfi Abdi, and Ali Douik. "Facial expression recognition via deep learning." In 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA), IEEE, (2017): 745-750.

[10] Perveen, Nazia, Nazir Ahmad, M. A. Q. B. Khan, Rizwan Khalid, and Salman Qadri. "Facial expression recognition through machine learning." International Journal of Scientific & Technology Research 5, no. 03 (2016).