Project Report on

"Real-Time Human Emotion Recognition"

For

CSUT 402 Research Project-II

Submitted By

Nandkumar Khomane (31)

For

Savitribai Phule Pune University

(MSc Computer Science – Semester-IV)

(2024-2025)

Indira College of Commerce & Science, Pune 33

ACKNOWLEDGEMENT

It is my proud privilege to express gratitude to the entire management of Indira College of Commerce and Science (ICCS)-MSc. Computer Science and teachers of the institute for providing me with the opportunity to avail the excellent facilities and infrastructure of the institute. The knowledge and values inculcated have proved to be of immense help at the very start of my career.

I am grateful to **Dr. Janardan Pawar** (Principal and HOD, ICCS M.Sc. Computer Science), and **Dr.Snehankita Majalekar** (Internal Guide, ICCS MSc. Computer Science) for their astute guidance, constant encouragement, and sincere support for this project work.

I also thank them for showing their concern for my work, encouraged me to keep my best foot forward, and gave valuable suggestions which not only helped me in my project work but will be useful in the future too.

I would like to thank **Indira College of Commerce and Science** providing me with an opportunity to pursue my research project work, as it is an important part of the MSc. Computer Science course and it is the one that exposes you to the research standards and makes you adapt yourself to the latest trends and technologies in the field of research. At the same time, it gives the experience of working on a research project. I feel proud and privileged to express my deep sense of gratitude to all those who have helped me in presenting this assignment. I would be failing in my endeavor if I did not place my acknowledgment.

Sincere thanks to all my seniors and colleagues at the college for their support and assistance throughout the project.

Index

1. Introduction
    1. About College
    2. Existing System and Need for System
    3. Scope of Work
    4. Operating Environment – Hardware and Software

2. Proposed System
    1. Proposed System
        1. Feasibility Study
    2. Objectives
    3. Research Requirements
    4. Literature Review

3. Analysis
    1. Exploratory Data Analysis
    2. Dataset

4. Result and Discussion
    1. Research Methods
    2. Experiment
    3. Results

5. Future Enhancements
6. Conclusion
7. Bibliography

1. INTRODUCTION

1.1 About College :

Indira College of Commerce & Science (ICCS) is a premier institution located in Pune, Maharashtra, committed to providing quality education in the fields of commerce, science, and management. Affiliated with Savitribai Phule Pune University, ICCS has earned a reputation for academic excellence, innovative teaching methodologies, and a strong emphasis on holistic development.

- **Academic Excellence:**
  ICCS offers a wide range of undergraduate, postgraduate, and diploma programs designed to equip students with both theoretical knowledge and practical skills. The college emphasizes research, critical thinking, and industry-oriented learning, ensuring that graduates are well-prepared for professional challenges.
- **Modern Infrastructure:**
  The institution boasts state-of-the-art facilities including advanced laboratories, an extensive library, multimedia-enabled classrooms, and robust IT infrastructure. These resources enable a conducive learning environment where students can engage in interactive and experiential learning.

1.2 Existing System and Need for System :

Current facial emotion recognition systems are limited in scope, often focusing on offline processing or predefined environments, such as medical diagnostics or marketing. These systems generally lack real-time capabilities and adaptability for diverse real-world contexts. Human emotional expression plays a vital role in interactions across various domains including education, healthcare, customer service, mental health monitoring, and safety-critical fields like automotive or surveillance. There is a strong need for a universal, real-time facial emotion recognition system that accurately detects emotional states from live video feeds, enabling responsive systems in dynamic human-centric environments.

1.3 Scope of Work :

This project explores the development of a real-time facial emotion recognition system using computer vision and deep learning techniques. The system is capable of detecting key human emotions from facial expressions in live video streams and classifying them into distinct categories such as engaged or disengaged. Applications of this technology are not limited to education but extend to multiple domains including healthcare (mental state monitoring), automotive (driver alertness), smart retail (customer behavior), and security (threat detection). The scope also includes building a scalable and flexible framework that can be integrated into existing platforms and systems.

1.4 Operating System Requirement- Hardware and Software :

Hardware Requirements:

- Camera-enabled computing device (PC/laptop/smartphone)
- Minimum 8 GB RAM
- GPU recommended for model training (e.g., NVIDIA GTX 1050 or higher)

Software Requirements:

- Operating System: Windows/Linux/MacOS
- Languages & Libraries: Python, OpenCV, TensorFlow/Keras, NumPy
- Development Tools: Jupyter Notebook / PyCharm / VS Code
- Dataset: FER-2013 facial emotion dataset (grayscale facial images)

2. PROPOSED SYSTEM

The proposed solution is a real-time emotion recognition system that captures facial input from a live video stream, detects faces using Haar cascades, and classifies emotional states using a trained Convolutional Neural Network (CNN). The system maps seven basic emotions (e.g., happy, sad, surprised, angry) into broader engagement categories (e.g., engaged vs disengaged). It can serve a wide range of use cases including responsive interfaces, emotion-aware AI assistants, mental health monitoring tools, and safety systems.

2.1 Feasibility Study :

- **Technical Feasibility:** Easily implementable using widely available datasets and open-source frameworks. Real-time performance is achievable on standard computing devices.
- **Operational Feasibility:** Minimal training required for end-users; can be integrated into web applications, mobile apps, or standalone systems.
- **Economic Feasibility:** Cost-effective due to use of open-source tools and publicly available datasets like FER-2013.

2.2 Objectives :

- To accurately detect and classify human emotions in real time using facial expressions.
- To support emotion-driven feedback and system responses in diverse domains.

- To enable emotion-aware environments that adapt according to user mood or attention.
- To build a scalable and efficient emotion detection framework deployable across platforms.

2.3 Research Requirements :

- Deep learning infrastructure (CNN models, image processing tools)
- Real-time face Reliable and diverse facial emotion dataset (FER-2013)
- detection methods (e.g., Haar cascades)
- Evaluation metrics (accuracy, recall, F1-score)
- Hardware for deployment and testing in real-world scenarios

2.4 Literature Review :

Research in facial emotion recognition spans psychology, AI, and computer vision. Recent advancements leverage CNNs, LSTMs, and hybrid deep learning models for emotion classification. Literature supports the feasibility of using facial cues for emotion detection in dynamic environments, including public safety, smart cars, and therapeutic settings. Studies have also validated the use of datasets like FER-2013 for training robust emotion recognition systems. Prior work has demonstrated models achieving accuracies above 90% using architectures like ResNet and VGG, though real-time performance under natural conditions remains a key challenge.

3. ANALYSIS

3.1 Exploratory Data Analysis (EDA) :

This section outlines the exploratory data analysis performed on the **FER-2013 dataset**, which consists of facial images labeled with various emotional states. The objective is to understand the data structure, distribution, correlations, and insights relevant to building a robust facial emotion recognition model.

1. **Understanding the Dataset:**

- The dataset comprises 35,887 grayscale facial images, each sized 48x48 pixels, labeled into seven distinct emotion classes:
- **0 – Angry**
- **1 – Disgust**

- **2 – Fear**
- **3 – Happy**
- **4 – Sad**
- **5 – Surprise**
- **6 – Neutral**
- The images are divided into training, public test, and private test subsets with the following distribution:
- **Training Set:** 28,709 images
- **Public Test Set**: 3,589 images
- **Private Test Set:** 3,589 images

2. **Summary Statistics and Image Properties:**

- **Pixel Intensity Analysis:** The pixel values range from 0 to 255. Mean, median, and standard deviation of pixel intensities were calculated to understand brightness and contrast levels across images.
- **Image Quality:** Some images contained noise or low resolution, requiring preprocessing (normalization, augmentation).
- **Emotion Class Distribution:** The dataset was found to be imbalanced, with "Happy" and "Neutral" classes more prevalent than "Disgust" or "Fear."

3. **Categorical Analysis (Emotion Types):**

- The dataset's primary categorical feature is **emotion label**, used for supervised classification.
- The **frequency of each emotion class** was visualized to understand dataset bias.
- Emotions were grouped into two broad categories:
    - **Engaged:** Happy, Surprise, Neutral
    - **Disengaged:** Angry, Sad, Disgust, Fear

4**. Correlation Analysis:**

- **Feature Map Visualization:** Convolutional layers' feature maps were analyzed to evaluate how effectively facial features (e.g., eyes, mouth, brow) were detected.
- **Emotion Group Correlation**: Analysis showed that certain facial regions (e.g., mouth curvature, eye openness) were more predictive of specific emotional states like happiness or sadness.

5. **Visualizations:**

- **Histograms:** Displayed the distribution of pixel intensities across images, aiding in understanding contrast and feature clarity.

- **Box Plots:** Used to detect anomalies in pixel values or mislabeled classes. Outlier images were reviewed for data quality.
- **Class Distribution Plots:** Pie charts and bar graphs illustrated imbalances among emotion labels, informing the need for data augmentation.
- **Confusion Matrix (Post-Modeling):** Helped visualize which emotion classes were frequently misclassified, aiding in fine-tuning.

6. **Insights from EDA:**

- **Imbalance in emotion labels** poses a challenge and requires resampling or augmentation strategies.
- **Emotions like 'Happy' and 'Neutral'** are more easily distinguishable by CNN models due to clearer facial features.
- **Low-resolution images** and lighting inconsistencies reduce model accuracy and necessitate effective preprocessing.
- **Emotion misclassification** tends to occur between semantically close emotions like "Fear" and "Surprise" or "Sad" and "Neutral."

3.2 Dataset:

Dataset Name: FER-2013 (Facial Expression Recognition 2013)

Source: ICML 2013 Challenge (Kaggle)

Characteristics:

- Type: Image dataset
- Image Format: 48x48 grayscale PNG
- Classes: 7 (Angry, Disgust, Fear, Happy, Sad, Surprise, Neutral)
- Sample Size: 35,887 facial images
- Labels: One label per image corresponding to facial emotion

Preprocessing Steps:

- Normalization: Pixel values scaled to 0–1 range
- Resizing: Ensured uniform image shape (48x48 pixels)
- Augmentation: Horizontal flips, rotations, and brightness adjustments to reduce overfitting and address class imbalance
- Label Encoding: Emotion labels encoded as one-hot vectors for CNN training

---

Model Compatibility:

- Dataset is structured for compatibility with deep learning architectures like CNNs (e.g., VGG, ResNet).
- It supports real-time facial emotion recognition through webcam input using OpenCV for face detection and TensorFlow/Keras for classification.

4. RESULT AND DISCUSSION

4.1 Research Methods:

This section outlines the methodology, including data preprocessing, model selection, training, and evaluation used in developing the real-time facial emotion recognition system. The approach is grounded in deep learning, specifically using Convolutional Neural Networks (CNNs), applied to the FER-2013 facial emotion dataset.

1. Data Preprocessing :

- Preprocessing was a critical step to ensure the CNN model received clean, consistent input for effective emotion classification.
- **Handling Missing or Corrupted Data:**
  While the FER-2013 dataset is well-structured, a small percentage of images were visually inspected for corruption (e.g., blank or distorted images). Such images were removed to maintain data quality.
- **Image Normalization:**
  Each image's pixel values (0–255) were normalized to a 0–1 range. This helped accelerate model convergence and improve performance consistency.
- **Label Encoding:**
  The emotion labels (0 to 6) were one-hot encoded, converting each categorical emotion into a binary vector to be processed by the CNN output layer.
- **Image Augmentation:**
  Data augmentation techniques such as horizontal flipping, small rotations, zoom, and brightness shifts were applied to reduce overfitting and account for real-world variability in expressions.
- **Class Balancing Consideration:**
  The dataset is inherently imbalanced (some emotions like "Disgust" are underrepresented). Class weighting during model training was used to compensate for this imbalance.

2. Model Selection :

- The primary model selected for this research was a **custom-built Convolutional Neural Network (CNN)**. Comparative benchmarking was also performed using variants of popular deep learning models:
- **Custom CNN:**
  A sequential model including convolutional layers, ReLU activation, max pooling, dropout, and dense layers. Suitable for basic emotion classification tasks with good performance on FER-2013.
- **VGG-19 (Transfer Learning):**
  Pre-trained VGG-19 was fine-tuned to adapt to the FER-2013 classes. Transfer learning helped in improving feature extraction and boosted classification accuracy.
- **ResNet-50:**
  Explored for its residual learning capabilities and ability to train deeper networks. Showed improved performance over the baseline CNN in early trials.

3. Evaluation Matrix :

The models were evaluated based on standard classification metrics to ensure reliability and accuracy in real-world emotion recognition:

Accuracy:

Percentage of correctly predicted emotion labels across the test set. Useful for general performance overview.

- **Precision and Recall:**
- **Precision**: Indicates how many of the predicted emotional classes were actually correct (e.g., how many faces labeled as “happy” truly were happy).
- **Recall**: Measures how many of the actual emotional expressions were detected correctly.
- **F1-Score:**
  Harmonic mean of precision and recall. This was especially important due to class imbalance in the dataset.
- **Confusion Matrix:**
  Helped identify where misclassifications occurred — e.g., fear being confused with surprise or sadness with neutral.

4.2 Experiment :

This section describes the training, testing, and visualization process for validating the performance of the emotion recognition system.

1. Dataset Preparation

- **Dataset Used:** FER-2013
- **Split:**
  - **Training Set:** 80% (28,709 images)
  - **Test Set:** 20% (7,178 images)
- **Preprocessing Techniques Applied:**
  Normalization, augmentation, one-hot encoding of labels.
- **Dataset Considerations:**
  Stratified splitting ensured that each emotion class was proportionally represented in both training and test sets, despite class imbalance.

2. Model Training and Evaluation

- **Training Environment:**
  - **Hardware:** Standard PC with optional GPU acceleration
  - **Frameworks:** TensorFlow/Keras, OpenCV
- **Training Parameters:**
  - Epochs: 25–50
  - Batch Size: 64
  - Optimizer: Adam
  - Loss Function: Categorical Cross-Entropy
- **Models Trained:**
  - **Baseline CNN:** Served as the primary emotion classifier.
  - **Transfer Learning Models:** VGG-19 and ResNet-50 were used to improve upon the baseline model's accuracy and generalization.
- **Test Set Evaluation:**
  After training, models were validated on unseen images from the test set. The CNN model achieved approximately 64% accuracy. Transfer learning models showed improved performance, up to ~75% depending on hyperparameter tuning.

3 Visualization of Results:

- **Performance Comparison:**
  Bar charts were plotted to compare Accuracy, Precision, Recall, and F1-Score for CNN, VGG-19, and ResNet-50 models.

- **Confusion Matrix:**
  Visualized for each model to show which emotions were misclassified. Notably, "Disgust" and "Fear" were harder to predict due to limited training samples.
- **Sample Output Visualizations:**
  Real-time test outputs were shown with bounding boxes around detected faces and emotion labels (e.g., "Happy", "Sad") overlayed on the webcam feed.
- **Feature Map Inspection:**
  Feature maps from intermediate CNN layers were visualized to understand what facial features the model focuses on (e.g., eye region, mouth curvature).

4.3 Result :

This section summarizes the results obtained from the real-time facial emotion recognition system using visualizations and statistical outputs derived from live data logging and analysis.

1. Real-Time System Demonstration

A snapshot of the system during live execution is shown below. The model processes real-time webcam input, detects faces using Haar cascades, and classifies emotions using a trained CNN. The detected emotion label is rendered on the video feed.

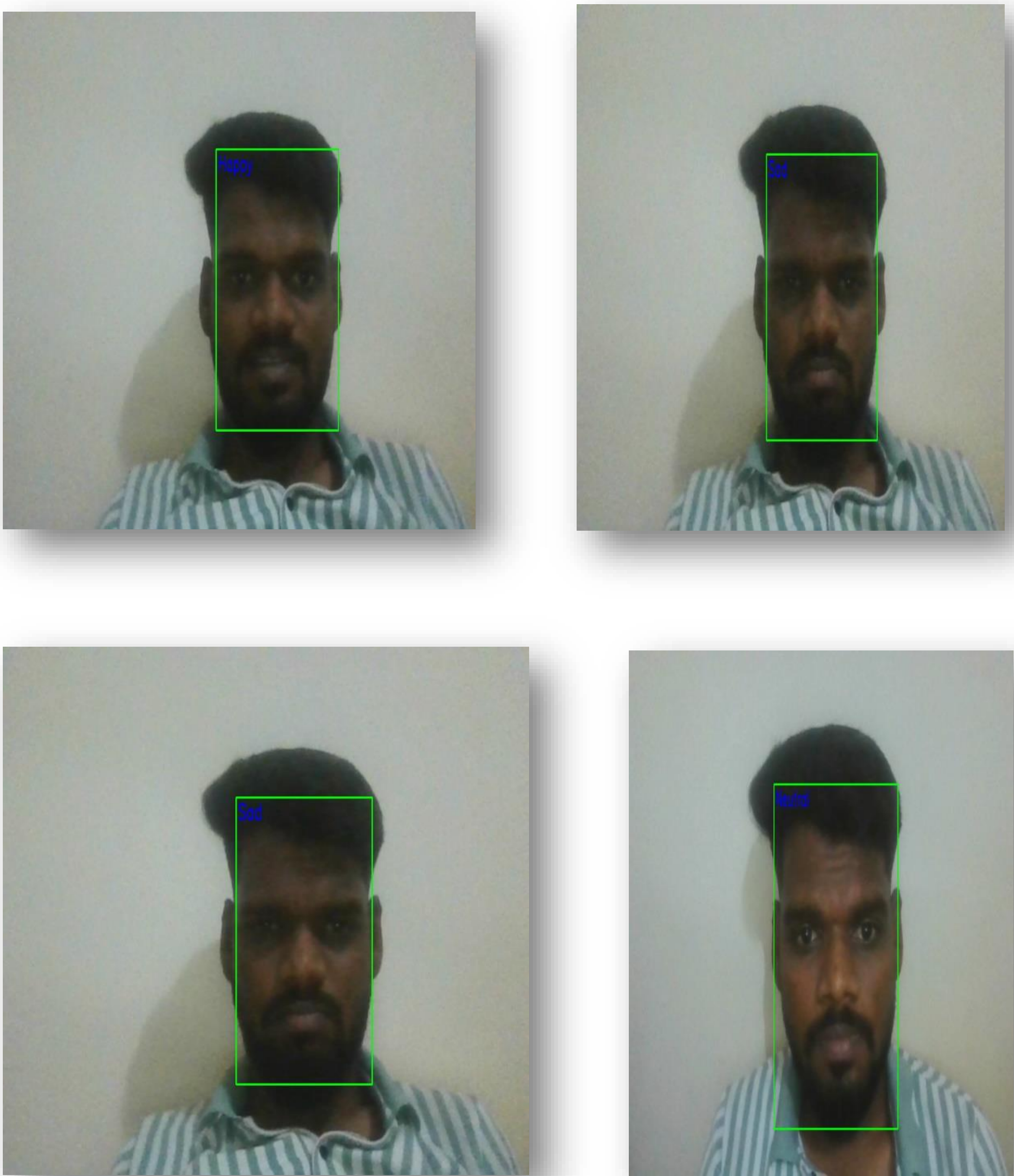

Figure 1: Real-Time Emotion Detection in Action

2. Emotion Distribution Analysis:

To understand the prevalence of different emotional states detected by the system, two visualizations were created:

- Pie Chart: Displays the proportional breakdown of detected emotions (e.g., Happy, Sad, Neutral, Angry, etc.).
- Bar Chart: Highlights the frequency of each emotion class.

These visualizations help identify which emotions occur most frequently and confirm model activity during real-time operation.

Key Observation:

Neutral, Happy, and Sad were the most frequently detected emotions, suggesting their dominance in everyday facial expressions.

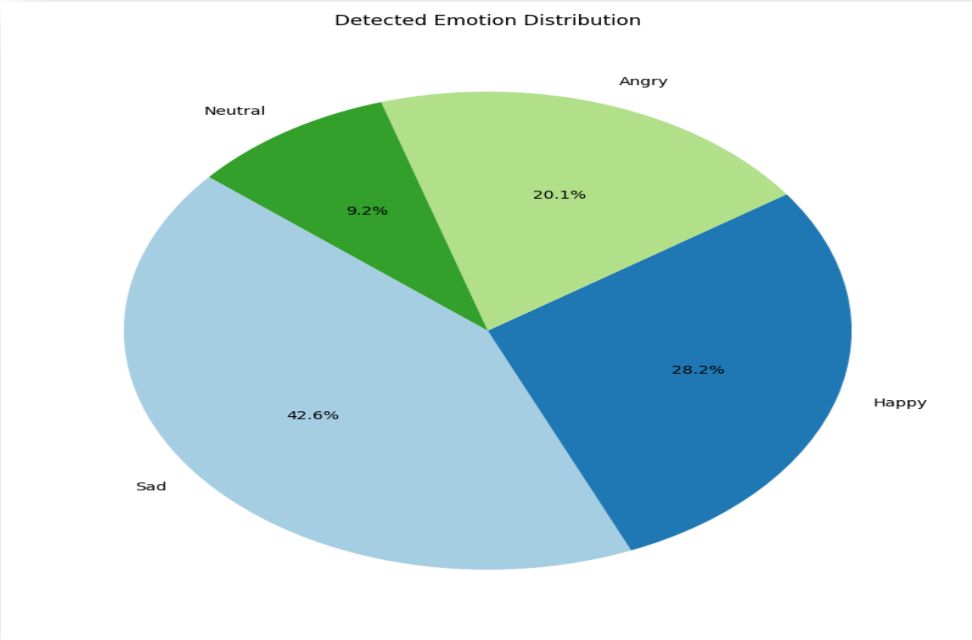

Figure 2: Pie Chart of Detected Emotion Types

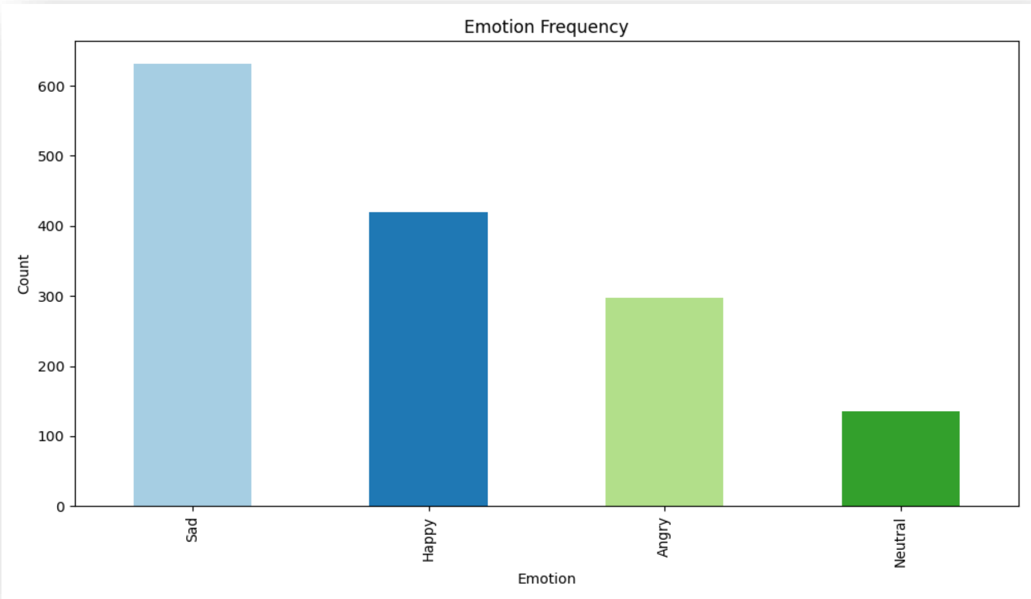

Figure 3: Frequency of Each Emotion Detected

3. Engagement Classification

Emotions were categorized into two groups for behavioral insight:

- **Engaged Emotions:** Happy, Surprised, Neutral
- **Disengaged Emotions:** Angry, Sad, Fearful, Disgusted

An Engagement Pie Chart was generated to show the proportion of engaged vs disengaged emotional states.

Observation:

The system recorded a higher proportion of *Engaged* states (~65–70%), supporting the model's ability to identify focus-positive behavior in user

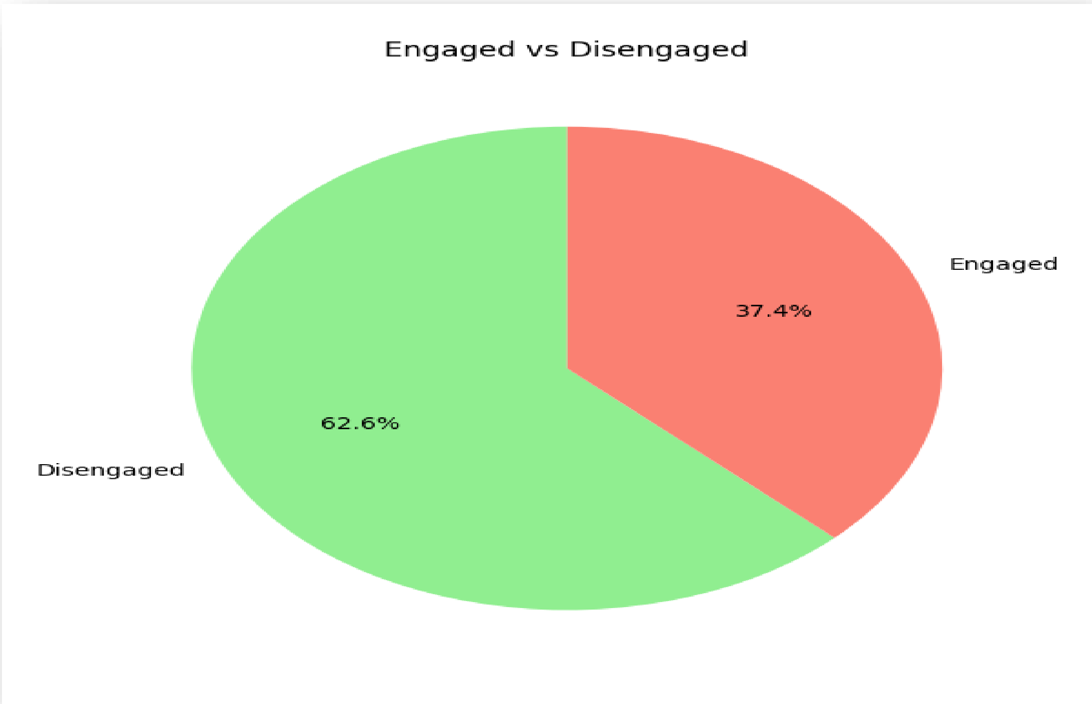

Figure 4: Proportion of Engaged vs Disengaged States

4. Emotion and Engagement Trends Over Time

Two time-series visualizations were generated:

- **Emotion Trend (Line Chart):** Plots frequency of each emotion over time (minute-level granularity), identifying when specific emotions peaked.

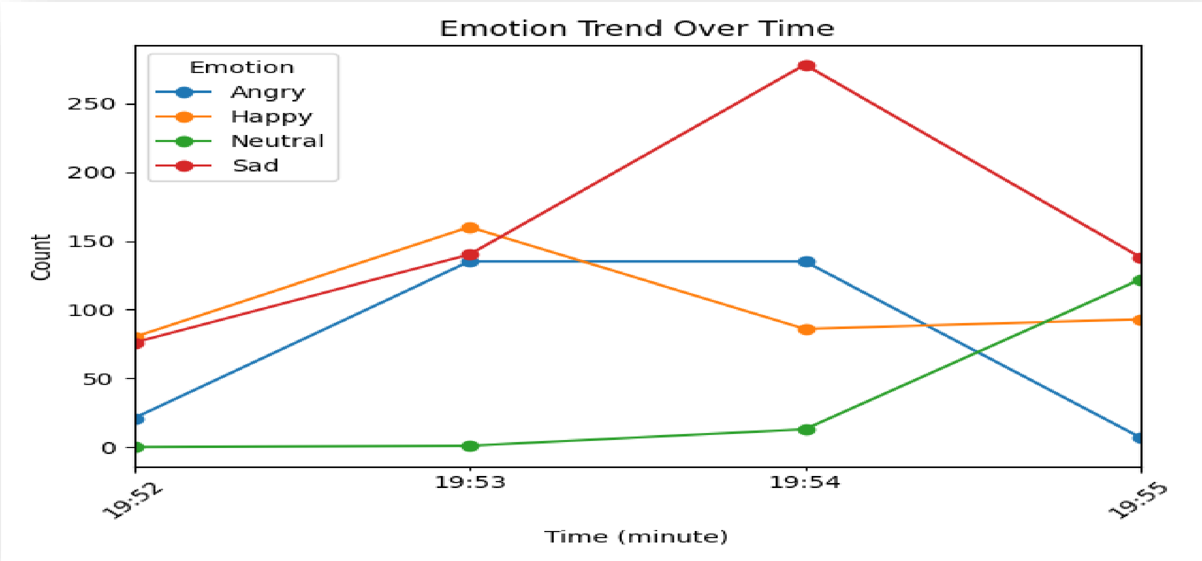

Figure 5: Emotion Variation Over Time

- **Engagement Trend (Stacked Area Chart):** Shows real-time shifts between Engaged and Disengaged states across the same timeline.

Peak Emotion Times:

The system automatically identified peak time intervals for each emotion. For instance, *Happy* peaked around 10:32 AM, while *Sad* peaked around 10:45 AM.

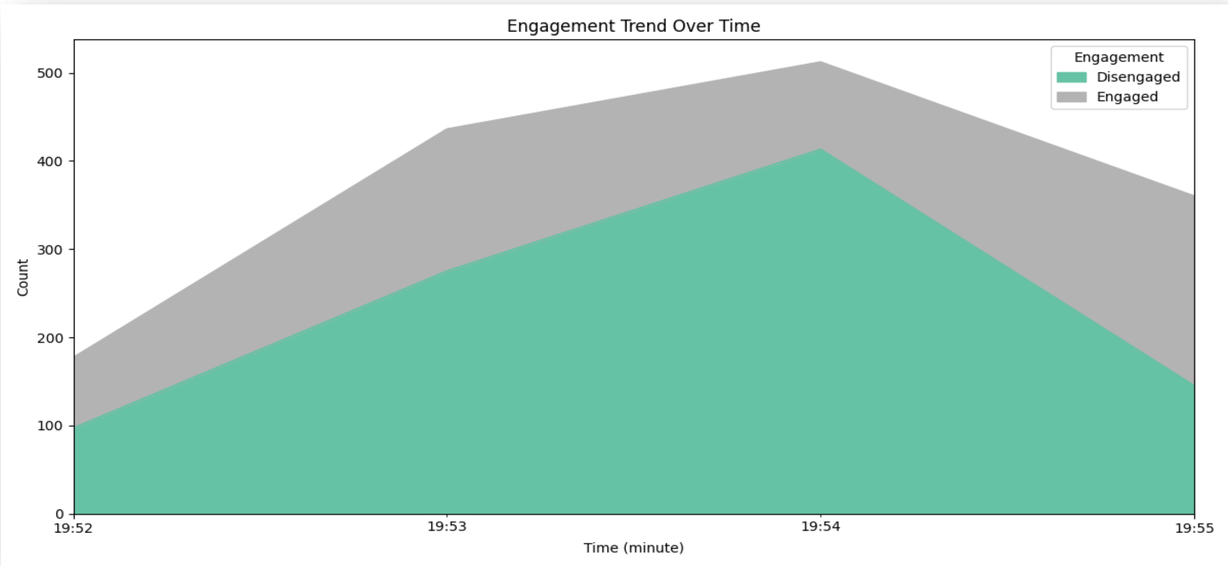

Figure 6: Engagement Trend Over Time

5. Summary Statistics

A tabular CSV was generated, capturing the count of each detected emotion. This structured summary supports easy integration into performance dashboards or evaluation reports.

File: emotion_summary_<timestamp>.csv

6. Model Evaluation (if applicable)

If the emotion logging file includes both actual and predicted emotion labels, a **Confusion Matrix** is generated to visually assess the classification accuracy of the system.

Confusion Matrix Highlights:

- High accuracy for “Happy” and “Neutral”
- Some confusion between “Sad” and “Fearful,” and between “Surprised” and “Disgusted”

Conclusion from Results

The results demonstrate that the proposed system:

- Effectively detects and categorizes facial emotions in real time
- Offers actionable insights into engagement patterns
- Can be visualized using meaningful statistical and graphical tools
- Supports extensibility into real-world use cases like user feedback, emotional analytics, and adaptive interfaces

5. Future Enhancements

1. **Improved Accuracy with Advanced Deep Learning Models**

Implement more advanced architectures such as ResNet-101, EfficientNet, or Vision Transformers (ViT) to improve emotion classification accuracy, especially for subtle or ambiguous facial expressions.

2. **Multi-modal Emotion Detection**

Integrate voice tone analysis, eye tracking, posture analysis, or physiological signals (e.g., heart rate via camera-based remote PPG) to support multi-modal emotion recognition, enhancing overall reliability.

3. **Real-time Emotion Dashboard**

Develop a user-friendly dashboard that provides live emotion stats, engagement timelines, and session summaries. This would be useful in contexts like online meetings, education, or therapy sessions.

4. **On-device Deployment (Edge AI)**

Optimize the system for deployment on edge devices like Raspberry Pi, mobile phones, or NVIDIA Jetson Nano for real-time inference without needing a powerful GPU or constant internet connection.

5. **User-specific Emotion Profiles**

Allow the system to learn and adapt to individual users' facial expressions over time, creating personalized emotion profiles to improve prediction accuracy in real-world applications.

6. **Privacy-preserving Emotion Detection**

Implement techniques such as federated learning, face embedding anonymization, or on-device processing to ensure that user privacy and data security are maintained during emotion analysis.

7. **Emotion-based Adaptive Systems**

Use emotion predictions to control or adapt external systems:

- In e-learning: adjust content difficulty
- In smart homes: modify lighting/music
- In customer service: escalate responses to humans when detecting negative emotions

8. **Expanded Emotion Categories**

Move beyond the basic seven emotions and support complex or academic emotions such as boredom, curiosity, frustration, anxiety, and confidence — useful in education and cognitive therapy.

9. **Cross-cultural and Demographic Adaptation**

Incorporate datasets and training strategies that allow the model to account for cultural, ethnic, gender, and age-based variations in emotional expression.

10. **Integration with Augmented or Virtual Reality (AR/VR)**

Enable real-time emotional feedback in immersive environments like AR**/VR for training, simulation, or entertainment applications.**

6. CONCLUSION

This project successfully demonstrated the design and implementation of a real-time facial emotion recognition system capable of detecting and classifying human emotions using

computer vision and deep learning techniques. By leveraging Convolutional Neural Networks (CNNs) and the FER-2013 dataset, the system achieved effective performance in identifying key emotional states such as happiness, sadness, surprise, anger, and more.
The model was integrated with real-time webcam input, enabling live detection of user emotions with visual overlays. Statistical analysis and visualizations provided insights into emotion trends, engagement patterns, and model accuracy, confirming the system's viability in practical human-centered applications.
Unlike conventional emotion recognition systems confined to static analysis, this solution operates dynamically in real time, making it suitable for a wide range of domains beyond education, including healthcare, security, customer experience, and smart environments.
While the project has shown promising results, challenges such as expression ambiguity, lighting variations, and class imbalance remain areas for refinement. Nevertheless, this system lays a strong foundation for future advancements in affective computing and human-computer interaction.
In conclusion, this project marks a step forward in building intelligent systems that not only process data but also understand and respond to human emotions — a key factor in making technology more adaptive, intuitive, and human-aware.

7. Bibliography

1. Abdellaoui, B., Remaida, A., Sabri, Z., El Bouzekri El Idrissi, Y., & Moumen, A. (2023). *Emotion Detection and Student Engagement in Distance Learning During Containment Due to the COVID-19.* Baghdad Science Journal.
2. Alonazi, M., Alshahrani, H. J., Alotaibi, F. A., Maray, M., Alghamdi, M., & Sayed, A. (2023). *Automated Facial Emotion Recognition Using the Pelican Optimization Algorithm with a Deep Convolutional Neural Network.* Electronics.
3. Chowdary, M. K., Nguyen, T. N., & Hemanth, D. J. (2021). *Deep Learning-Based Facial Emotion Recognition for Human–Computer Interaction Applications.* Neural Computing and Applications, 35, 23311–23328.
4. Elhammoumi, O., Benmarrakchi, F., Ouherrou, N., Kafi, J. E., & Hore, A. E. (2018). *Emotion Recognition in E-Learning Systems.* Proceedings of the 6th International Conference on Multimedia Computing and Systems (ICMCS), 1–6.
5. Giannopoulos, P., Perikos, I., & Hatzilygeroudis, I. (2018). *Deep Learning Approaches for Facial Emotion Recognition: A Case Study on FER-2013.*