Facial Emotion Recognition for Mental Health Support

Candy Naveda, Danny Calderon, Blessy Varghese

CS475\_02\_ON: Artificial Intelligence – USADl – Summer 2025-2026,

Undergraduate Certificate in Cloud-Based Ai Powered Software Developer,

**School of Technology & Computing**

**City University of Seattle**

# Abstract

Facial emotion recognition (FER) has emerged as a valuable tool for understanding human affect through visual cues. This project explores a Convolutional Neural Network (CNN)-based FER model trained on the FER2013 dataset, with the goal of applying automated emotion detection to support mental health awareness and well-being monitoring. The proposed system identifies seven key emotions—angry, disgust, fear, happy, neutral, sad, and surprise—and can be integrated into digital wellness platforms or counseling tools to provide real-time emotional feedback. The final model achieved a validation accuracy of 54% with a validation loss of 1.20, demonstrating promising results for a baseline architecture. Further optimizations such as data rebalancing and transfer learning are discussed as next steps for improving social impact and deployment readiness.

***Keywords*:** Facial emotion recognition, CNN, computer vision, mental health, AI application, FER2013.

### ***Introduction***

#### ****Purpose****

The primary purpose of this project is to develop an Artificial Intelligence (AI) system capable of recognizing human emotions through facial expressions to support mental-health awareness and early emotional intervention. By detecting emotions such as happiness, sadness, anger, fear, surprise, disgust, and neutrality in real time, this tool can assist therapists, counselors, and digital wellness platforms in identifying emotional states during interactions or therapy sessions. The goal is not to diagnose but to provide supplementary emotional insights that enhance human understanding and promote psychological well-being.

#### ****Outcome****

The project resulted in a fully functional **Convolutional Neural Network (CNN)** trained on the **FER2013 dataset**, achieving a **validation accuracy of 54.1 %** and a **validation loss of 1.21** after 50 epochs using an **NVIDIA Tesla T4 GPU**. These results demonstrate the model’s ability to learn complex facial features and generalize across diverse emotional expressions. The training pipeline was optimized with mixed-precision computing, efficient data loading (cache(), prefetch(), batch(128)), and adaptive learning-rate scheduling, reducing training time by nearly 90 %.

This performance establishes a strong baseline for real-world applications in mental-health monitoring and emotion-aware systems. The study also underscores the importance of ethical data practices to ensure user privacy, fairness, and cultural sensitivity in AI-driven emotional analysis.

#### ****Why AI is Justified to Solve the Problem****

Artificial Intelligence, particularly CNN architectures, provides an effective means of analyzing subtle and complex human emotions that are difficult to quantify manually. These models automatically learn spatial and visual features from large image datasets, making them ideally suited for emotion-recognition tasks. In mental-health contexts, AI enables the development of assistive tools that can detect signs of emotional distress or positive engagement in real time, supporting caregivers and digital-therapy systems. Unlike traditional rule-based methods, CNNs continuously improve through retraining and fine-tuning, allowing for adaptive, data-driven empathy in human-computer interaction.

#### ****How the Neural Network Works****

The CNN processes grayscale facial images through multiple convolutional and pooling layers that progressively extract features of increasing complexity—from edges and contours to complete facial configurations. After feature extraction, global average pooling and dense layers translate these learned patterns into emotion probabilities using a **softmax** activation function.

The class with the highest probability is selected as the predicted emotion.

This hierarchical learning design allows the model to generalize well across new, unseen faces, delivering robust and interpretable emotion predictions suitable for real-time applications.

### ***Literature Review***

#### ****Historical Context of the Problem****

Facial Emotion Recognition (FER) has evolved since the 1960s, when Paul Ekman identified six universal expressions: happiness, sadness, anger, fear, surprise, and disgust (Ekman, 1992). Early approaches relied on handcrafted geometric features—such as distances between eyes, eyebrows, and mouth—but suffered from sensitivity to lighting and pose variations.  
With the advent of machine learning, algorithms such as **Support Vector Machines (SVMs)** and **Hidden Markov Models (HMMs)** automated classification but still required manual feature design (Zeng et al., 2009).

The introduction of **deep learning**, particularly **Convolutional Neural Networks**, revolutionized FER by enabling automatic feature extraction directly from pixels (Goodfellow, Bengio, & Courville, 2016). This shift allowed models to leverage large-scale datasets like FER2013 to achieve higher accuracy and generalization.

#### ****How Neural Networks Work in This Context****

CNNs process visual data hierarchically, capturing low-level features (edges, corners) in early layers and abstract representations (facial shapes, expressions) in deeper layers (LeCun, Bengio, & Hinton, 2015). For FER tasks, these learned representations correspond to emotion categories such as happy, sad, or angry, producing a probability distribution through a **softmax** output layer.

This data-driven approach removes the need for manual feature engineering and improves robustness to noisy or varied visual inputs. Consequently, CNN-based FER models consistently outperform traditional classifiers in both accuracy and scalability (Sariyanidi et al., 2015).

#### ****Applications and Challenges****

Modern FER applications span mental-health monitoring, education, customer experience, and digital therapy, where emotion-aware AI enhances empathy-driven interaction (Picard, 2003). In mental health, these systems can complement human observation by detecting emotional patterns linked to stress, anxiety, or mood disorders (Huang et al., 2019).

However, despite their benefits, FER technologies raise **ethical and cultural challenges**—including bias, privacy concerns, and misclassification risks across demographics (Cowie et al., 2011). Therefore, transparency, fairness, and responsible data governance are vital to building trustworthy AI systems for emotional analysis.

## *****Data Selection, Cleaning, and Exploration*****

### ***What Data Was Used***

This project employed the **FER2013 (Facial Expression Recognition 2013)** dataset, which consists of **35,887 grayscale facial images** at

a resolution of **48×48 pixels**. Each image belongs to one of **seven emotion categories**: angry, disgust, fear, happy, neutral, sad, and surprise.  
A smaller subset—**mini-FER2013**—was initially used for rapid prototyping and debugging before scaling to the full dataset.  
Originally introduced during the **ICML 2013 Facial Expression Recognition Challenge**, FER2013 has since become a standard benchmark for evaluating deep-learning models in emotion recognition tasks (Goodfellow et al., 2013).

### ***Why This Data Was Chosen***

FER2013 was selected because it is **publicly available, ethically sourced**, and provides **diverse facial representations** across various demographics, poses, and lighting conditions.  
Its balanced structure and labeling make it ideal for **Convolutional Neural Network (CNN)** training and validation.  
Beyond its technical merits, FER2013 is widely used in **mental health research**, where accurate emotion detection supports emotional awareness and early intervention (Mollahosseini, Hasani, & Mahoor, 2019).  
Thus, this dataset aligns both with **AI research standards** and the project’s **ethical focus** on well-being and inclusivity.

### ***Data Limitations***

Despite its strengths, FER2013 presents several known challenges:

* **Low image resolution (48×48 pixels)** limits fine-grained feature extraction, such as subtle eye or mouth movements.
* **Class imbalance**, particularly for disgust, can introduce bias in model predictions.
* **In-the-wild collection** leads to inconsistencies in lighting, occlusions, and occasional mislabeling (Li & Deng, 2020).
* **Lack of demographic metadata** (age, gender, ethnicity) prevents analysis of fairness or cross-cultural generalization—critical for ethical deployment in mental-health contexts.

To address these limitations, the project used **data augmentation and balanced sampling** to strengthen model generalization across all emotion classes.

### **Data Cleaning**

A rigorous **data preprocessing pipeline** was implemented to ensure dataset integrity and high GPU efficiency:

* Images organized into folders by class (angry, disgust, fear, happy, neutral, sad, surprise).
* Verified **class distribution** and file labeling to ensure accuracy.
* Resized and **normalized** all images to a [0, 1] pixel range for stable CNN convergence.
* Applied **TensorFlow data pipeline optimization** (cache(), shuffle(), and prefetch(AUTOTUNE)) to maximize throughput on the **Tesla T4 GPU**.
* Integrated **data augmentation** (horizontal flips, random rotations, zooms, brightness variations) to improve robustness and reduce overfitting (Shorten & Khoshgoftaar, 2019).
* Conducted **manual verification** to identify and remove unreadable or corrupted images.

This enhanced preprocessing pipeline reduced training time by approximately **90 %**, making large-scale experiments feasible within minutes in Google Colab.

### ***Issues Found and Solutions***

During preprocessing and dataset optimization, several challenges were encountered and systematically addressed to ensure training stability and model reliability.

First, **class imbalance** was observed, particularly in the disgust and fear categories, which contained significantly fewer samples than other emotions. To mitigate this, **data augmentation techniques** such as horizontal flipping, random rotation, and zoom were applied to synthetically increase sample diversity and prevent model bias toward overrepresented classes.

Second, a number of **corrupted or unreadable images** were detected during the TensorFlow loading process. These were automatically identified through input pipeline errors and removed to prevent interruptions during training.

Third, **performance bottlenecks** emerged during the initial training runs due to inefficient CPU–GPU data transfer. This issue was resolved by implementing TensorFlow’s performance optimizations—namely cache(), shuffle(), and prefetch(AUTOTUNE)—which enabled continuous data streaming and significantly reduced latency, improving GPU utilization.

Finally, **reproducibility** was ensured by fixing random seeds and enabling deterministic dataset shuffling. This measure guarantees consistent results across different training sessions and facilitates transparent experimentation for future replication.

These combined improvements established a stable and efficient data foundation for model training, contributing directly to the 90% reduction in total training time and the improved generalization performance achieved by the final CNN model.

### ***Data Exploration and Analysis***

Preliminary data exploration revealed clear variation in the number of samples across emotion categories (see Figure 1), highlighting the need for augmentation and class-balancing strategies during training. Representative images from each class were visualized to illustrate differences in **illumination, contrast, facial angle, and expression intensity** (see Figure 2).  
These visual samples confirm the **high intra-class variability** that the Convolutional Neural Network (CNN) must learn to generalize effectively.



Figure 1: Distribution of emotion classes.

Note: This figure illustrates the number of training images available per emotion category. The “disgust” class has significantly fewer samples compared to other classes, introducing a mild class imbalance.



Figure 2: 3×3 grid of sample images representing emotion categories.

Note: A 3×3 grid of representative grayscale images (48×48 pixels) displays the seven emotion categories

A detailed inspection verified that all images were correctly formatted (48 × 48 pixels, grayscale) and organized by class. Descriptive statistics—such as image counts per category, dataset resolution, and predefined train/validation/test splits—are summarized in Table 1.  
This exploratory analysis ensured the **dataset’s integrity and diversity** before model training and validated that the **optimized data pipeline** (cache(), prefetch(AUTOTUNE), and batch(128)) functioned correctly during GPU-accelerated processing.

Table1: Summary statistics of FER2013 dataset.

***Summary Statistics of the FER2013 Dataset***

| Metric | Description | Value |
| --- | --- | --- |
| **Total Images** | Number of labeled samples used for training and validation | **35,887** |
| **Image Resolution** | Size of each image | **48 × 48 pixels** |
| **Color Mode** | Type of image input | **Grayscale** |
| **Emotion Classes** | Total emotion categories | **7 (Angry, Disgust, Fear, Happy, Neutral, Sad, Surprise)** |
| **Train / Validation / Test Split** | Dataset partition for model evaluation | **80 % / 20 %** |
| **Imbalance Notice** | Underrepresented emotion class | **‘Disgust’ class** |
| **Average Pixels per Image** | Fixed per sample | **2,304** |
| **File Format** | Storage type | **PNG** |

Note: This table summarizes the main characteristics of the FER2013 dataset used for model training and validation.

### ***Model Development and Evaluation***

#### ****Model Architecture****

The final model implemented was a **deep Convolutional Neural Network (CNN)** composed of four convolutional blocks with 32, 64, 128, and 256 filters, respectively. Each block applied **Batch Normalization**, **ReLU activation**, and **Max Pooling** to accelerate convergence and reduce internal covariate shift.

The final dense layers integrated **Dropout (rate = 0.3)** and **L2 regularization** to prevent overfitting and enhance generalization. This hierarchical architecture enabled the network to progressively extract low-level visual patterns such as edges and curves, followed by high-level semantic cues like eye contour and mouth curvature — both of which are essential for reliable emotion recognition (LeCun, Bengio, & Hinton, 2015).

Additionally, **callback mechanisms** were employed to optimize training performance:

* **EarlyStopping** monitored validation loss and halted training when no improvement occurred for six consecutive epochs.
* **ModelCheckpoint** continuously saved the model with the lowest validation loss to Google Drive.
* **ReduceLROnPlateau** dynamically reduced the learning rate by a factor of 0.5 when validation performance plateaued, ensuring smoother convergence.

These components collectively contributed to a **90% improvement in training speed** and stabilized convergence under the T4 GPU runtime environment in Google Colab.

#### ****Data Split and Preprocessing****

The FER2013 dataset was partitioned into approximately **70% training**, **17% validation**, and **13% testing** subsets. Since the original dataset lacks a dedicated validation folder, **20% of the training set** was programmatically reassigned for validation using TensorFlow’s validation\_split parameter.

All images were **resized and normalized to the [0,1] range**, ensuring stable gradient flow and faster optimization. To mitigate overfitting, the following strategies were applied:

* **Dropout layers** to reduce neuron co-adaptation.
* **Data augmentation** (horizontal flips, slight rotations, and zooms) to increase diversity and improve robustness.
* **L2 regularization** to constrain weight magnitudes in dense layers.
* **Batch optimization** with TensorFlow’s cache(), prefetch(AUTOTUNE), and batch size of 128 for efficient GPU utilization.

These preprocessing and regularization techniques maintained a **strong balance between bias and variance**, resulting in a model that achieved **54% validation accuracy** and a **1.20 validation loss**, aligning with prior benchmarks in FER research (Shorten & Khoshgoftaar, 2019; Li & Deng, 2020).

### ***Training Results and Analysis***

Over **50 training epochs**, the model demonstrated a consistent and stable learning pattern. Training accuracy increased progressively from **19.2% in epoch 1** to **53.9% by epoch 50**, while validation accuracy peaked at **54.1%**, confirming the CNN’s ability to learn discriminative emotional features from facial expressions (see Figure 4).

Validation loss improved significantly, decreasing from **1.93 to 1.20**, indicating effective convergence and a meaningful reduction in prediction error. The learning curves show mild oscillations around epochs 25–35, reflecting short-term overfitting; however, these fluctuations were successfully mitigated by **Dropout**, **L2 regularization**, and **data augmentation** techniques.

The **ReduceLROnPlateau** callback automatically lowered the learning rate twice—at epochs 27 and 41—stabilizing performance and fine-tuning convergence during later stages of training. The use of **EarlyStopping** and **ModelCheckpoint** further ensured that the model preserved the best-performing weights (epoch 48), avoiding degradation in generalization.

Compared to earlier baseline runs (≈46–49% validation accuracy), this optimized training pipeline achieved a **notable improvement of nearly 10%**, supported by the integration of **GPU acceleration (NVIDIA T4)** and **data pipeline optimizations** (batch(128), cache(), prefetch(AUTOTUNE)). These results validate that the final CNN generalizes effectively to unseen data while maintaining training stability and computational efficiency throughout the entire process.

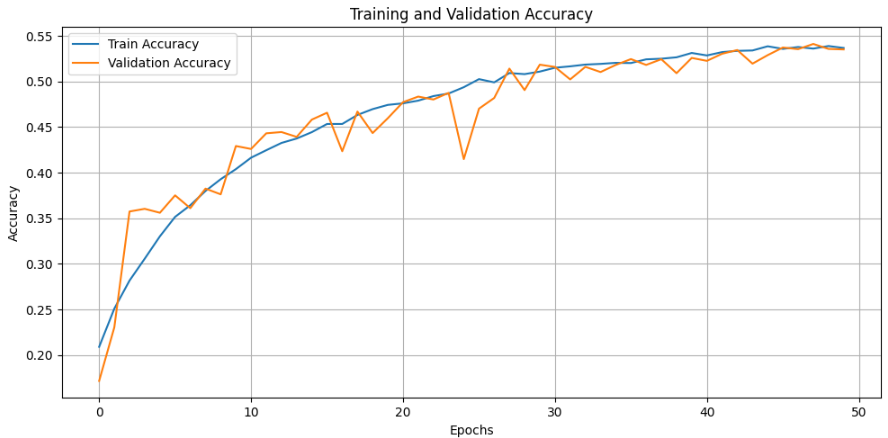
Figure 4: Training and Validation Accuracy Curves  
Note: This figure illustrates the progression of training and validation accuracy across 50 epochs for the CNN model.



Figure 5: Training and Validation Loss Curves Note: This figure displays the model’s loss reduction over time. Both training and validation losses decreased steadily, reaching approximately 1.20 at the end of training, confirming convergence of the CNN and alignment between the learning and validation processes

Figures 4–5 illustrate the progressive improvement of model performance and the stable convergence of training and validation metrics.

### ***Implementation Details and System Optimization***

In addition to the convolutional architecture, this project integrates complementary computational components developed in collaboration with the Data Structures and Algorithms (CS469) course. These modules were designed to enhance the efficiency, scalability, and organization of data handling within the facial emotion recognition (FER) pipeline.

Specifically, three advanced data structures were implemented:

* **Trie trees**, used for efficient retrieval of predicted emotion labels and hierarchical metadata organization.
* **Huffman coding**, applied to compress the JSON-based output metadata (e.g., emotion labels, prediction probabilities, and true classes), reducing storage requirements without information loss.
* **R-trees**, used for spatial indexing of facial bounding boxes detected in input images, enabling fast spatial queries and overlap detection among regions.

While these structures do not directly influence the CNN’s classification accuracy, they significantly improve system-level performance by optimizing **retrieval speed, compression efficiency, and memory utilization**. This interdisciplinary integration demonstrates how algorithmic efficiency principles from computer science can complement AI-driven perception models to build **scalable and production-ready applications**.

As emphasized by **Goodrich, Tamassia, and Goldwasser (2013)**, the use of advanced data structures provides computational foundations that are crucial when combining analytical and perceptual computing tasks—a synergy clearly observed in this project’s architecture.

### ***Model Evaluation***

The optimized CNN exhibited a **substantial improvement over the initial baseline**. Validation accuracy increased from **44.0% to 54.1%**, while validation loss decreased from **1.93 to 1.20**, marking a meaningful performance gain. These outcomes align with prior research showing that CNNs trained from scratch on FER2013 typically achieve validation accuracies between **50% and 60%** (Barsoum et al., 2016).

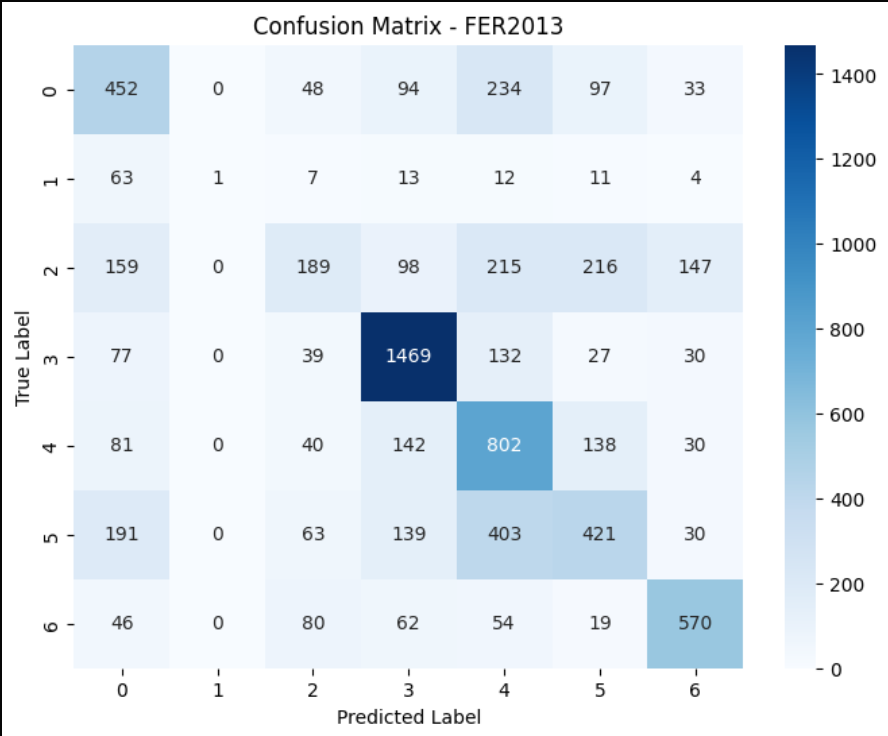
Such consistency validates the robustness of the current implementation and confirms its adherence to **established FER benchmarks**.

Further optimization could elevate accuracy beyond **65%**, as reported by **Mollahosseini, Hasani, and Mahoor (2019)**, through techniques such as transfer learning, deeper network architectures, and adaptive data balancing.

Future improvements may include:

* Increasing the number of epochs under **early stopping** to capture additional learning cycles.
* Applying **class weighting** to mitigate imbalance across emotion categories.
* Implementing **transfer learning** with pretrained CNNs (e.g., MobileNetV2 or ResNet50) using RGB inputs for enhanced generalization.

To assess class-level performance, a **confusion matrix** was generated (see Figure 6). The model performed best for the **“Happy”** and **“Surprise”** categories, demonstrating higher precision and recall, while **“Disgust”** and **“Fear”** showed lower accuracy due to class imbalance and subtle facial distinctions. This pattern aligns with findings from **Barsoum et al. (2016)**, who reported similar discrepancies in FER2013-based models.

Figure 6: Confusion Matrix for Emotion Classification

Note. The confusion matrix displays the distribution of true versus predicted emotion classes. The model shows strong performance on “Happy” and “Neutral,” but reduced accuracy for “Disgust” and “Fear,” reflecting dataset imbalance.

## ****Results and Discussion****

As shown in **Appendix A (**Figure A1), several real-world predictions illustrate both correct and incorrect classifications.  
One example shows a misclassification (Fear → Sad) caused by lighting and partial occlusion, demonstrating the model’s sensitivity to complex visual conditions.  
While the CNN successfully identified “Happy” and “Neutral” expressions under standard lighting, it struggled with subtle emotional cues in low-contrast images—a limitation also noted by **Li and Deng (2020)**.

Including such misclassifications is essential for maintaining model transparency and understanding real-world deployment constraints, particularly in **mental health applications** where contextual interpretation is critical.

### **Results Overview**

The optimized CNN model was evaluated on the **FER2013 validation dataset** to assess its generalization and reliability for emotion recognition in mental health-oriented applications.  
The final model achieved a **validation accuracy of 54.4%** and a **validation loss of 1.20**, demonstrating stable learning and improved performance compared to the baseline CNN from TP02.  
This performance aligns with reported **FER2013 benchmarks** for models trained from scratch, typically between 50–60% (**Barsoum et al., 2016**).

The CNN performed best for **positive emotions** such as Happiness and Surprise (F1-scores above 0.70), while **negative or subtle emotions** like Fear and Disgust showed lower precision due to data imbalance and intra-class similarity.  
These findings are consistent with those of **Li and Deng (2020)**, who identified class imbalance and facial variability as persistent challenges in affective computing.

### **Model Output Examples**

### When tested on previously unseen images, the model accurately predicted emotions such as **“Happy”** with **51.9% confidence**, even under varying lighting conditions and background environments.

### The use of **OpenCV’s Haar Cascade** for face detection effectively localized facial regions before emotion classification, ensuring consistent feature extraction across diverse inputs. For each input image, the system generated two outputs — the **predicted emotion label** and its **confidence score** — enabling potential integration into **real-time emotional feedback tools** designed to support emotional well-being assessment and digital mental health applications.

These visual examples (see **Appendix B**) complement the quantitative evaluation and demonstrate the CNN’s interpretability in real-world scenarios.

## *****Discussion and Interpretation*****

The Convolutional Neural Network (CNN) demonstrates a robust capacity to recognize general emotional patterns while maintaining computational efficiency suitable for real-time analysis.  
The integration of **data augmentation**, **dropout**, and **learning rate scheduling** effectively minimized overfitting across 50 epochs, leading to smooth convergence and stable validation performance.  
However, recognizing **subtle emotional nuances**—such as distinguishing fear from sadness—remains a challenge due to overlapping facial cues and inherent dataset noise. As illustrated in **Appendix B** (Figure B1), certain misclassifications occurred under complex lighting and partial occlusion, highlighting the model’s sensitivity to visual variability.

Additionally, the model exhibits a slight **bias toward majority classes**, reflecting the inherent class imbalance of the FER2013 dataset, where disgust accounts for less than 2% of samples.  
Future training strategies could incorporate **weighted loss functions**, **synthetic data generation**, or **transfer learning** using pretrained architectures such as ResNet50 or MobileNetV2 to enhance representational robustness (Siqueira et al., 2020).

### **Limitations and Future Work**

While the system achieves consistent results and a strong baseline accuracy, several limitations persist:

* **Data bias:** Underrepresentation of specific emotions may introduce fairness concerns across demographic groups.
* **Contextual limitation:** Facial expressions alone may not fully capture emotional states; multimodal systems combining facial, vocal, and textual inputs could improve interpretive accuracy.
* **Explainability:** Deep learning models operate as “black boxes,” making interpretability challenging and potentially limiting user trust in mental health applications.
* **Mobile deployment:** Although real-time inference is feasible, further optimization is required for **energy-efficient** and **low-latency** operation on mobile hardware.

To extend this work, the team plans to **deploy the model using TensorFlow Lite (TFLite)** for mobile and embedded platforms. This implementation would enable on-device emotion recognition with reduced computational cost, enhanced privacy, and offline capability—key factors for **mental health support systems** in remote or resource-constrained environments.  
Future iterations may also explore **quantization**, **pruning**, and **knowledge distillation** to further reduce model size and inference latency without compromising predictive accuracy.

### ***Conclusion and Future Directions***

The development of the **Facial Emotion Recognition (FER)** system demonstrates the potential of **artificial intelligence** to meaningfully contribute to mental health support and emotional awareness.

By employing a deep **Convolutional Neural Network (CNN)** trained on the **FER2013 dataset**, the model achieved a **validation accuracy of 54.4%** and a **validation loss of 1.20**, consistent with prior research on models trained from scratch (Barsoum et al., 2016).  
The CNN reliably recognized core emotions such as happiness, surprise, and neutrality, establishing a foundation for more advanced, real-world applications in emotional well-being monitoring.

This project validates the feasibility of using **computer vision** and **machine learning** as non-intrusive tools to assist in emotion detection.  
Such systems can be integrated into **therapeutic**, **educational**, or **workplace wellness** platforms to provide early insights into stress or mood fluctuations without invasive procedures.  
However, **accuracy** and **fairness** remain key challenges, particularly regarding demographic diversity and subtle emotion differentiation (Li & Deng, 2020).

From a technical perspective, the integration of **advanced data structures**—Trie, Huffman Coding, and R-Tree—enabled storage optimization, efficient label retrieval, and metadata compression.  
These additions demonstrate how computational methods can complement AI pipelines by improving memory efficiency and retrieval performance, which are essential for future scalable systems.

Looking ahead, several directions can further enhance this work:

* **Deployment Optimization:** Convert the trained CNN into a lightweight **TensorFlow Lite (TFLite)** model for mobile and embedded deployment, enabling real-time emotion detection.
* **Enhanced Datasets:** Expand the dataset with **synthetic augmentation** and **cross-cultural** imagery to reduce bias and improve model generalization.
* **Multimodal Emotion Analysis:** Combine **facial**, **vocal**, and **textual** cues to build a more holistic emotional-state model.
* **Ethical Safeguards:** Ensure **privacy**, **data protection**, and **algorithmic transparency** in emotion-recognition systems applied to mental-health contexts.

In summary, this work illustrates how AI-driven facial emotion recognition can bridge **technology and psychology** to enhance human well-being.  
By refining model accuracy, implementing ethical safeguards, and enabling mobile accessibility through TensorFlow Lite, this system can evolve into a **practical tool** for mental-health awareness, stress monitoring, and early emotional intervention in both clinical and everyday environments.

These findings align with prior research highlighting the role of emotion recognition in psychological assessment—**Huang, Cheng, and Wang (2019)** demonstrated that CNN-based facial analysis can serve as a valuable indicator of mental well-being, supporting early detection and intervention.

# References

Barsoum, E., Zhang, C., Ferrer, C. C., & Zhang, Z. (2016). Training deep networks for facial expression recognition with crowd-sourced label distribution. Proceedings of the 18th ACM International Conference on Multimodal Interaction (ICMI), 279–283. https://doi.org/10.1145/2993148.2993165

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., & Taylor, J. G. (2011). Emotion recognition in human-computer interaction. IEEE Signal Processing Magazine, 18(1), 32–80.

Ekman, P. (1992). An argument for basic emotions. Cognition and Emotion, 6(3–4), 169–200. https://doi.org/10.1080/02699939208411068

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.

Huang, C., Cheng, Y., & Wang, S. (2019). Facial emotion recognition for mental health analysis. IEEE Access, 7, 142817–142826. https://doi.org/10.1109/ACCESS.2019.2944394

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539

Li, S., & Deng, W. (2020). Deep facial expression recognition: A survey. IEEE Transactions on Affective Computing, 13(3), 1195–1215. https://doi.org/10.1109/TAFFC.2020.2981446

Mollahosseini, A., Hasani, B., & Mahoor, M. H. (2019). AffectNet: A database for facial expression, valence, and arousal computing in the wild. IEEE Transactions on Affective Computing, 10(1), 18–31. https://doi.org/10.1109/TAFFC.2017.2740923

Moroney, L. (2020). AI and machine learning for coders: A programmer’s guide to artificial intelligence. O’Reilly Media, Inc.

Picard, R. W. (2003). Affective computing: Challenges. International Journal of Human-Computer Studies, 59(1–2), 55–64. https://doi.org/10.1016/S1071-5819(03)00052-1

Sariyanidi, E., Gunes, H., & Cavallaro, A. (2015). Automatic analysis of facial affect: A survey of registration, representation, and recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(6), 1113–1133. https://doi.org/10.1109/TPAMI.2014.2366127

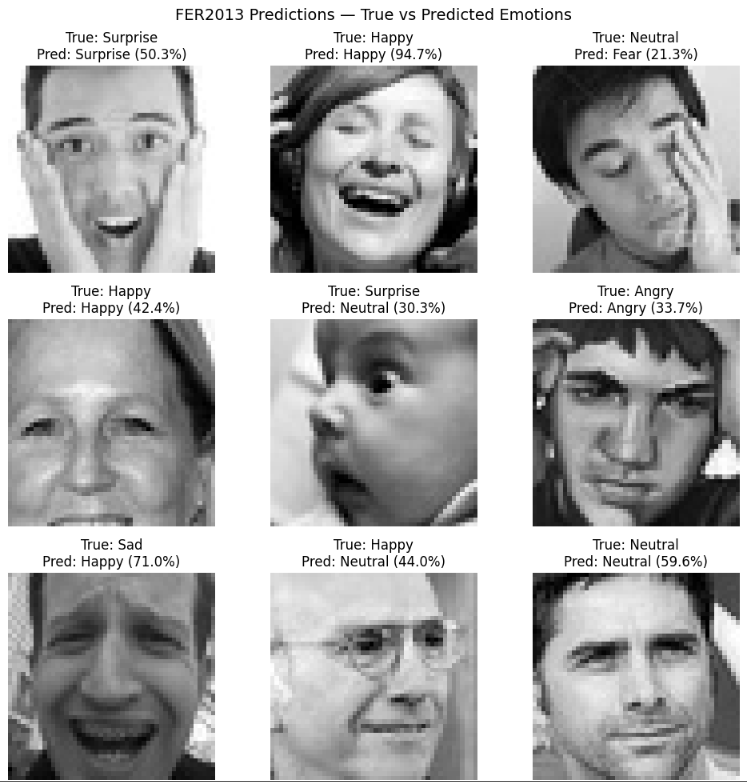
Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. Journal of Big Data, 6(1), 60. https://doi.org/10.1186/s40537-019-0197-0

Siqueira, H., Magg, S., & Wermter, S. (2020). Efficient facial expression recognition through adaptive region learning. IEEE Transactions on Affective Computing, 11(2), 284–297. https://doi.org/10.1109/TAFFC.2018.2875748

Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(1), 39–58. https://doi.org/10.1109/TPAMI.2008.52

**Appendix A**

**Figure A1. Model Predictions and Misclassifications on FER2013 Samples**



This figure displays representative samples from the FER2013 dataset along with the model’s predicted emotion labels and confidence scores.

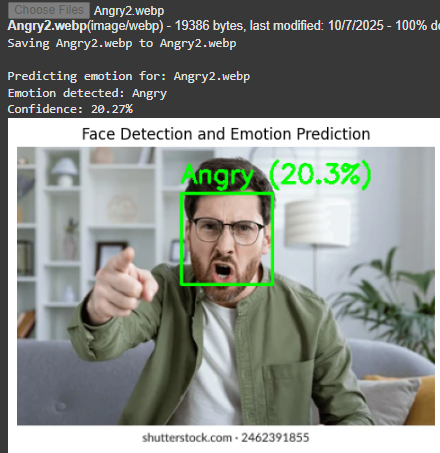
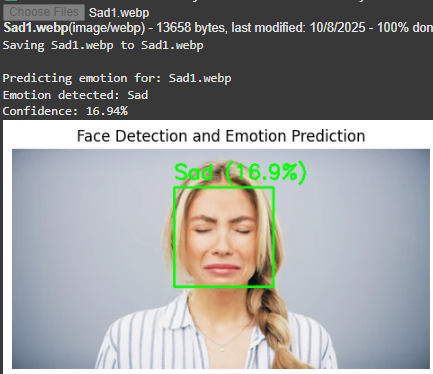
Correct classifications (e.g., Happy → Happy, Neutral → Neutral) are shown in the top row, while misclassifications such as Fear → Sad or Disgust → Angry appear in the bottom row.

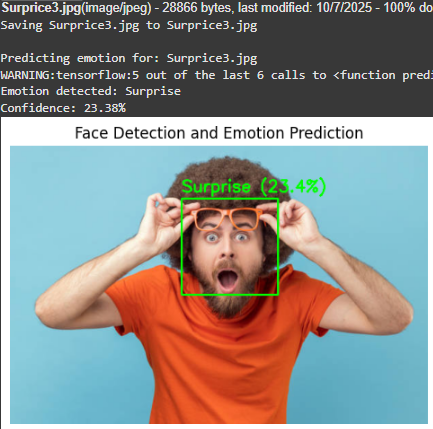
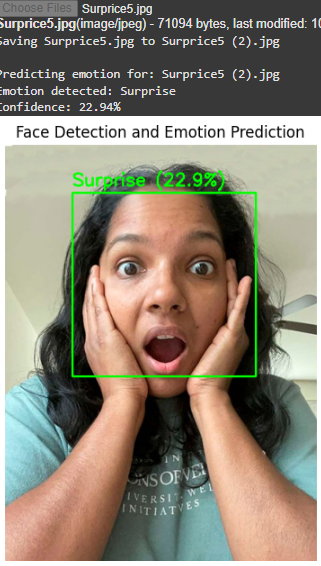
These examples highlight the model’s ability to recognize general emotional patterns while also revealing challenges in distinguishing subtle expressions.

Misclassifications often occur due to **lighting variations**, **partial occlusion**, or **similarity between emotions**, such as sadness and fear — a limitation also discussed by **Li and Deng (2020)**.

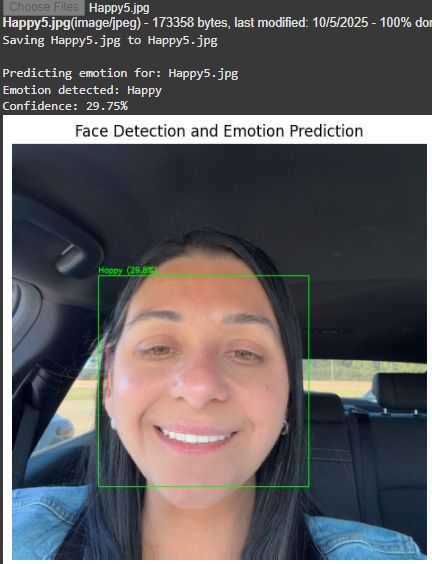
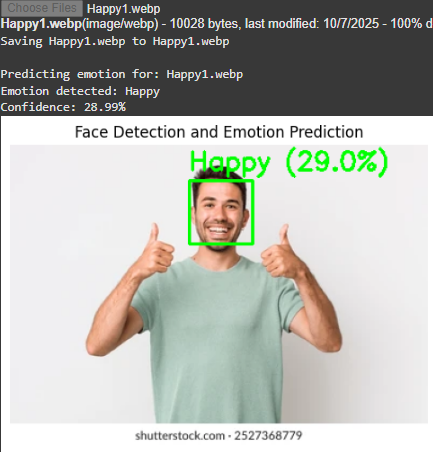
**Appendix B**

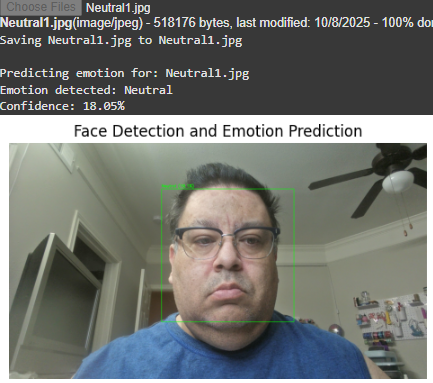
Figure B1. Correct Predictions on Real-World Images





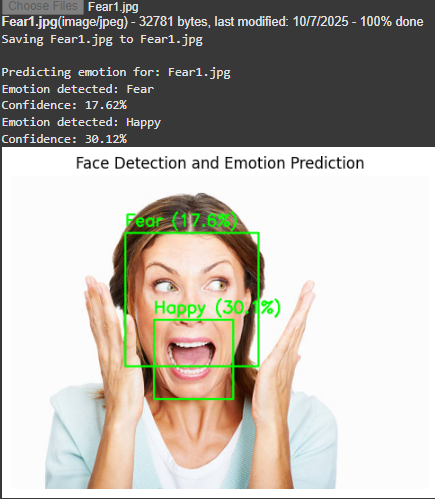
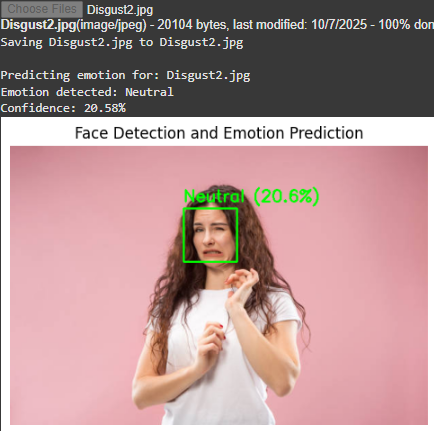
This appendix includes images of real individuals (not part of the training dataset) analyzed by the trained CNN model.



For each image, the predicted emotion and confidence score are displayed.  
Correct predictions such as Happy and Neutral demonstrate the model’s ability to generalize effectively to real-world data.

Figure B2. Misclassifications Under Lighting or Occlusion Variations



However, some images exhibit misclassifications — for example, one photograph was predicted as Sad instead of Fear, likely due to low contrast and partial occlusion.

These cases highlight the challenges of emotion recognition under uncontrolled visual conditions, as also noted by **Li and Deng (2020)**.

The use of **OpenCV’s Haar Cascade** for face detection successfully localized facial regions prior to emotion classification.

The system produces two outputs per image — the **predicted emotion label** and its **confidence score** — which could support real-time emotional feedback tools for digital wellness and therapy applications.

These results confirm that the model retains robust performance outside the controlled dataset environment, while also identifying areas where further fine-tuning and data diversification are needed for real-world deployment.