Facial Expression Recognition System Using Machine Learning

By: Waiyuk Kwong, Zhihui Chen, Tyler Lin, Blane R. York, and Carter D. Robinson

Introduction/Background

Facial expressions are a key aspect of human communication, conveying emotions and intentions without the need for words. As one of the most dynamic features of the human body, they provide critical information about a person's emotional state [2]. Recognizing these expressions through machine learning has applications in areas such as human-computer interaction, research, and education.

Literature Overview

> **Literature Overview**   ︿
>
> - **Paper 1 (Huang & Samonte, 2024)**: Explores the use of **Google MediaPipe** to track facial key points, combined with emotion analysis and other factors like eye movement, to assess engagement.
>
> - **Paper 2 (Sadikoğlu & Mohamed, 2022)**: Focuses on **Convolutional Neural Networks (CNNs)** and their ability to recognize facial expressions by extracting features from images. It highlights the advantages of CNNs and discusses the use of transfer learning.
>
> - **Paper 3 (Wang, Li, & Zhao, 2010)**: Describes the use of **Active Shape Models (ASM)** and **Support Vector Machines (SVM)** for real-time recognition, emphasizing geometric features for expression recognition.

Dataset Description

We will use the FER-2013 dataset, which contains over 35,000 labeled images across emotion categories like anger, fear, happiness, and sadness, to implement a model that can detect expressions in real time.

FER-2013 Dataset

Problem Definition

Problem:

- How can facial expressions be accurately recognized in real-time scenarios using facial detection systems?

Motivation:

- While current facial expression recognition systems are effective, they often struggle with real-time performance and robustness. By combining facial key points and image data, we aim to develop a more efficient and robust system.

Methods

Data Preprocessing Methods

1. **Facial Key Point Extraction**: Using **Google MediaPipe**, we will extract 468 facial key points from each image, capturing geometric changes that reflect subtle expressions [1]. MediaPipe is efficient for real-time key point detection and reduces the complexity of manually coding key point extraction.
   Library: `mediapipe`
2. **Image Normalization**: We will standardize image pixel values to reduce the effect of lighting inconsistencies and contrast variations, ensuring uniform input for the model.
   Library: `cv2`
3. **Data Augmentation**: Techniques like rotation, flipping, and cropping will be used to artificially expand the dataset and improve the model's ability to generalize across different facial angles and expressions.
   **Library:** `cv2` `numpy`

Machine Learning Algorithms/Models

1. **Random Forest**: Random Forest will use the facial landmark data to capture the unique patterns associated with different expressions using multiple decision trees. This approach provides a reliable model that's interpretable and relatively fast to train and evaluate.
2. **Support Vector Machine (SVM)**: SVM will use facial landmark data to classify expressions by finding the optimal boundaries between different classes based on landmark positions. This approach allows for high precision in recognizing expression.
3. **Convolutional Neural Network (CNN)**:The CNN will classify facial expressions directly from raw image data, using multiple layers to capture specific visual features and patterns that indicate different expressions. This layered approach enables the CNN to identify detailed facial features, making it effective at accurately recognizing expressions.

Results and Discussion

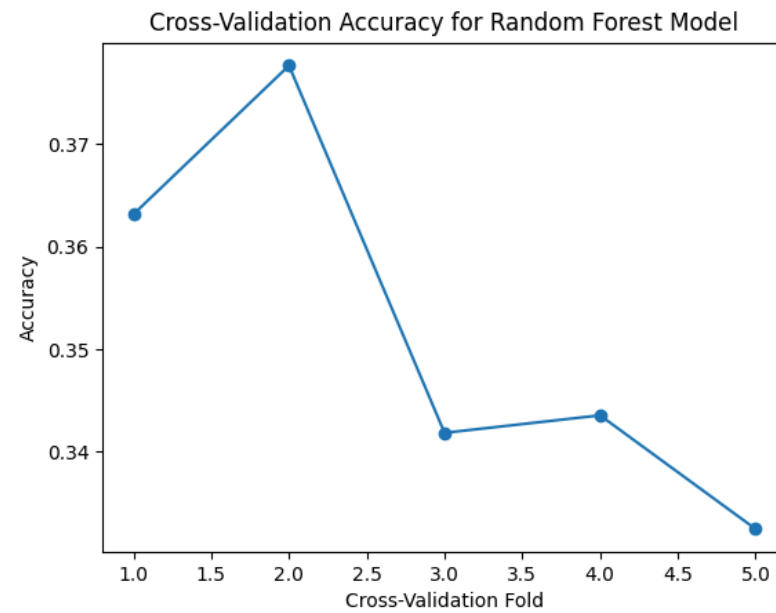## Random Forest Model

- **Accuracy:** 40.07%
- **F1-Score:** 0.37
- **Confusion Matrix:**
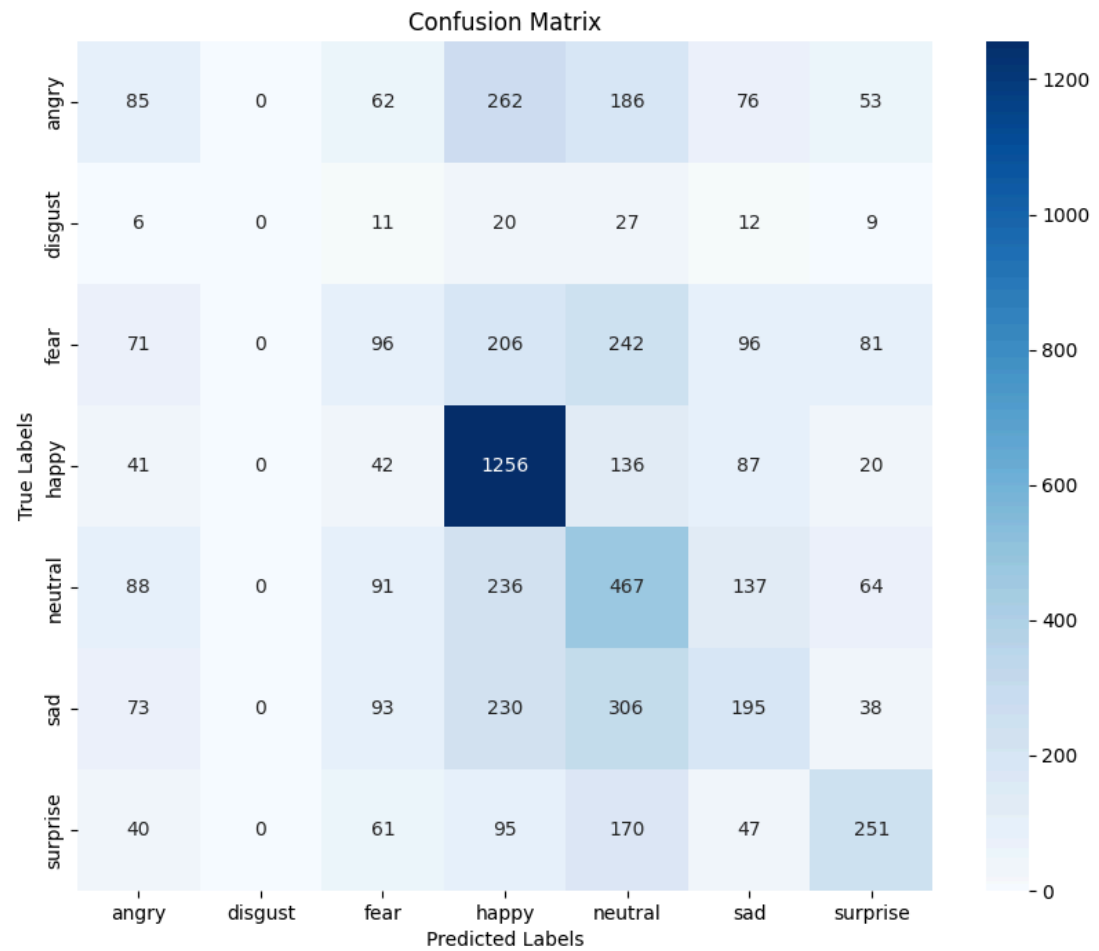
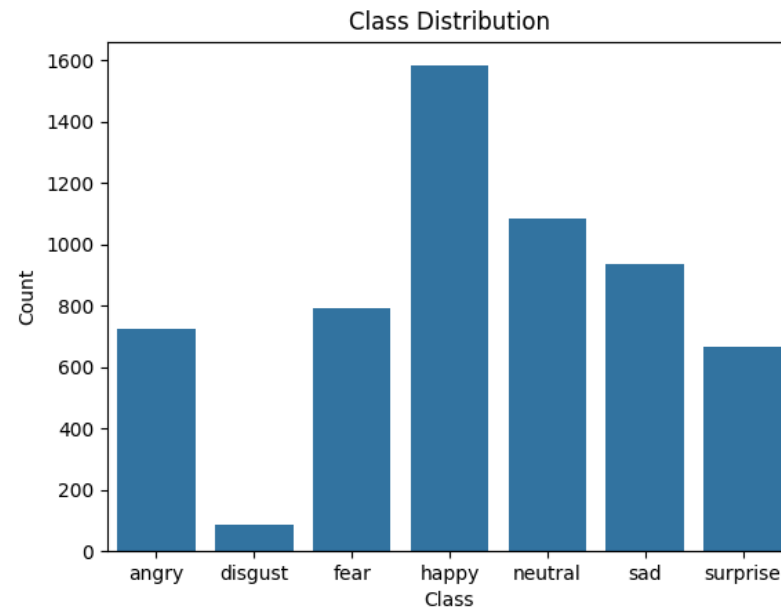## Confusion Matrix

Random Forest Confusion Matrix

- **Cross-Validation Accuracy:**

Cross-Validation Accuracy for Random Forest Model
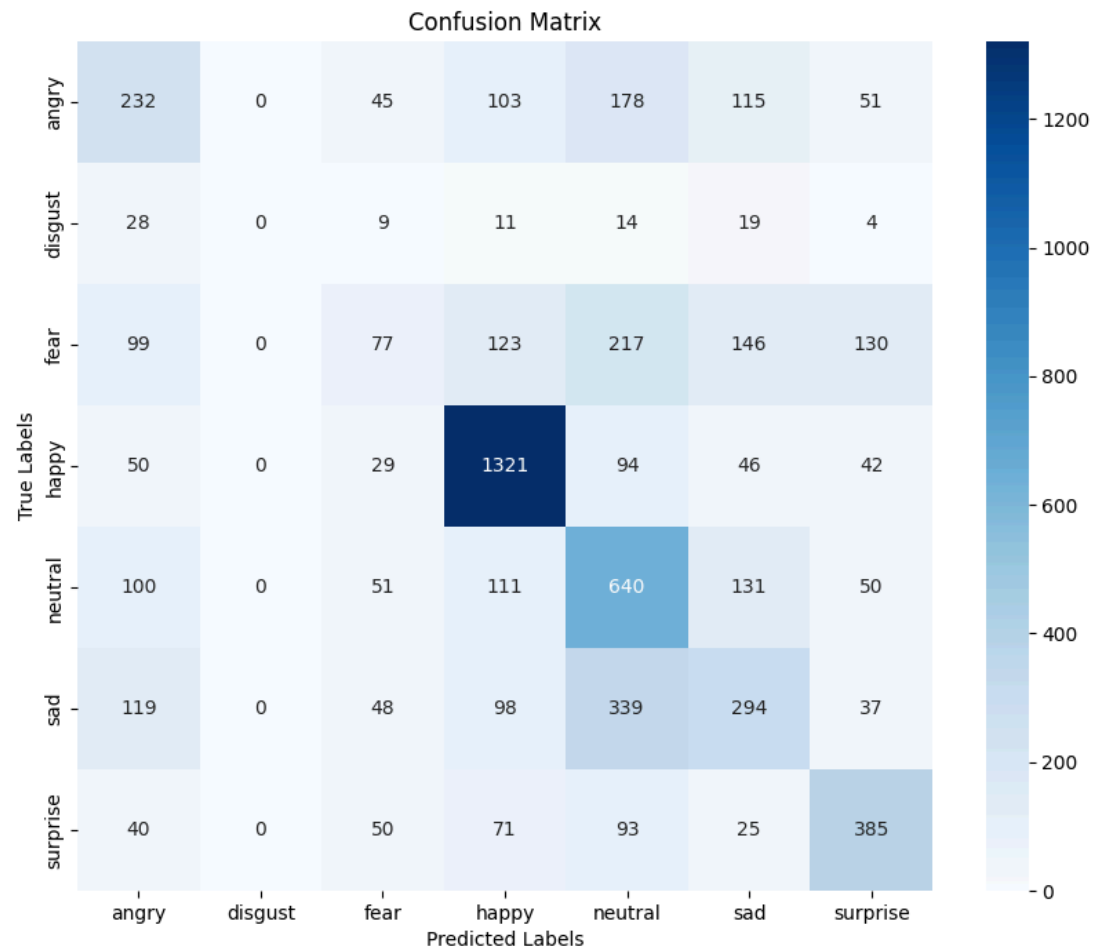
- **Class Distribution:**

Random Forest Analysis

The Random Forest model shows limited performance with an accuracy of 40.07% and an F1-score of 0.37. This indicates that the model struggles with generalization and class prediction. The confusion matrix and class distribution table show a strong bias toward predicting "happy" and a consistent failure to accurately predict "disgust." Additionally, the cross-validation accuracy plot shows that the model's performance is inconsistent across different data subsets, meaning its accuracy varies depending on the subset of data used for testing.
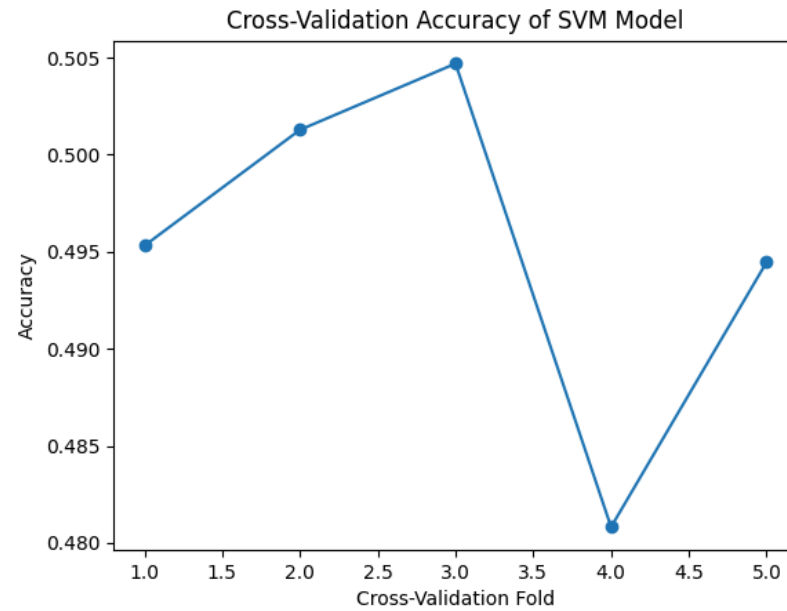
Support Vector Machine (SVM)

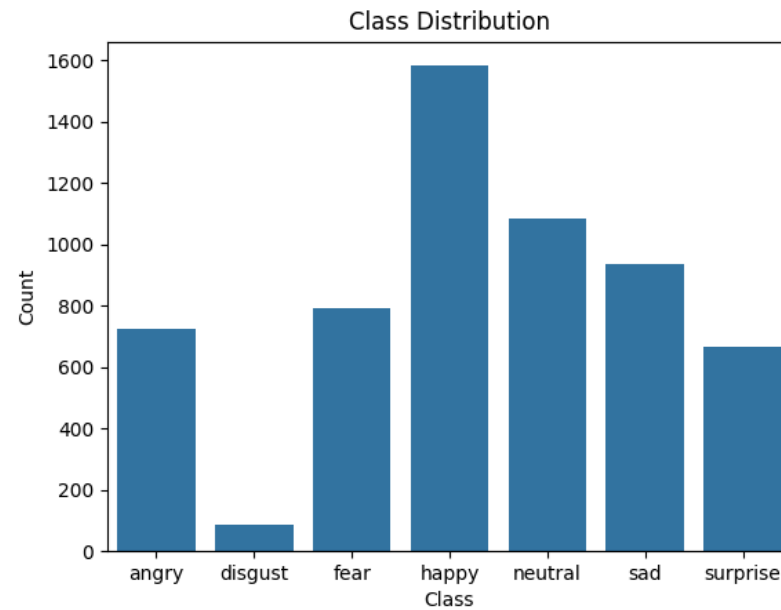- **Accuracy**: 50.28%
- **F1-Score**: 0.48
- **Confusion Matrix**:

Confusion Matrix

SVM Confusion Matrix

- **Cross-Validation Accuracy:**

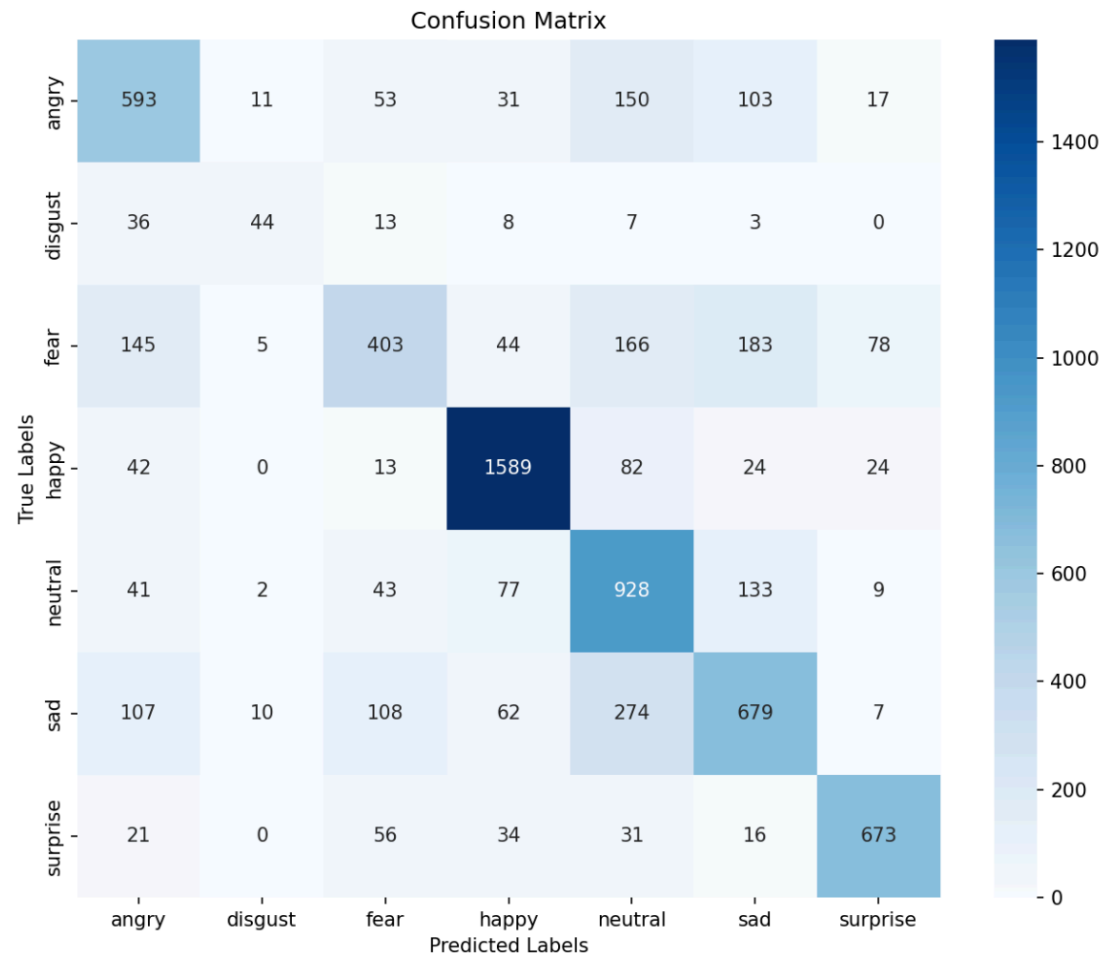Cross-Validation Accuracy for SVM

- **Class Distribution:**

Support Vector Machine Analysis

The Support Vector Machine Model ahieved an accuracy of 50.28% and F1-score of 0.48. The model performed better than the Random Forest Mode, but still struggles to classify facial expression acccurately. Similar to the Random Forest model, the SVM model is also biased toward predicting "happy" and misclassifyling other expressions. Although the overall accruacy in cross validation accuracy plot is around 50%, the variability across folds shows that the SVM model's performance also depends on the specific data subset used for training and testing.
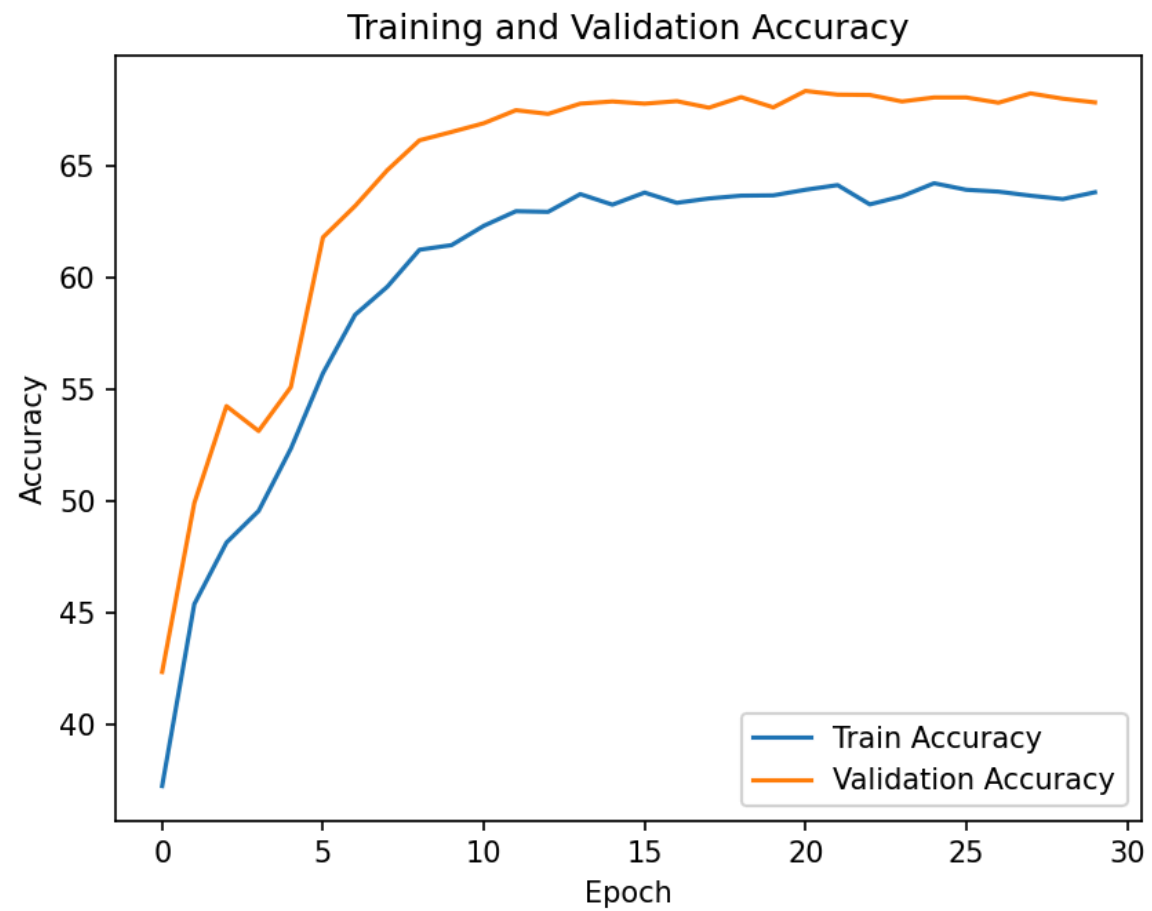
Improved Convolutional Neural Network (CNN)

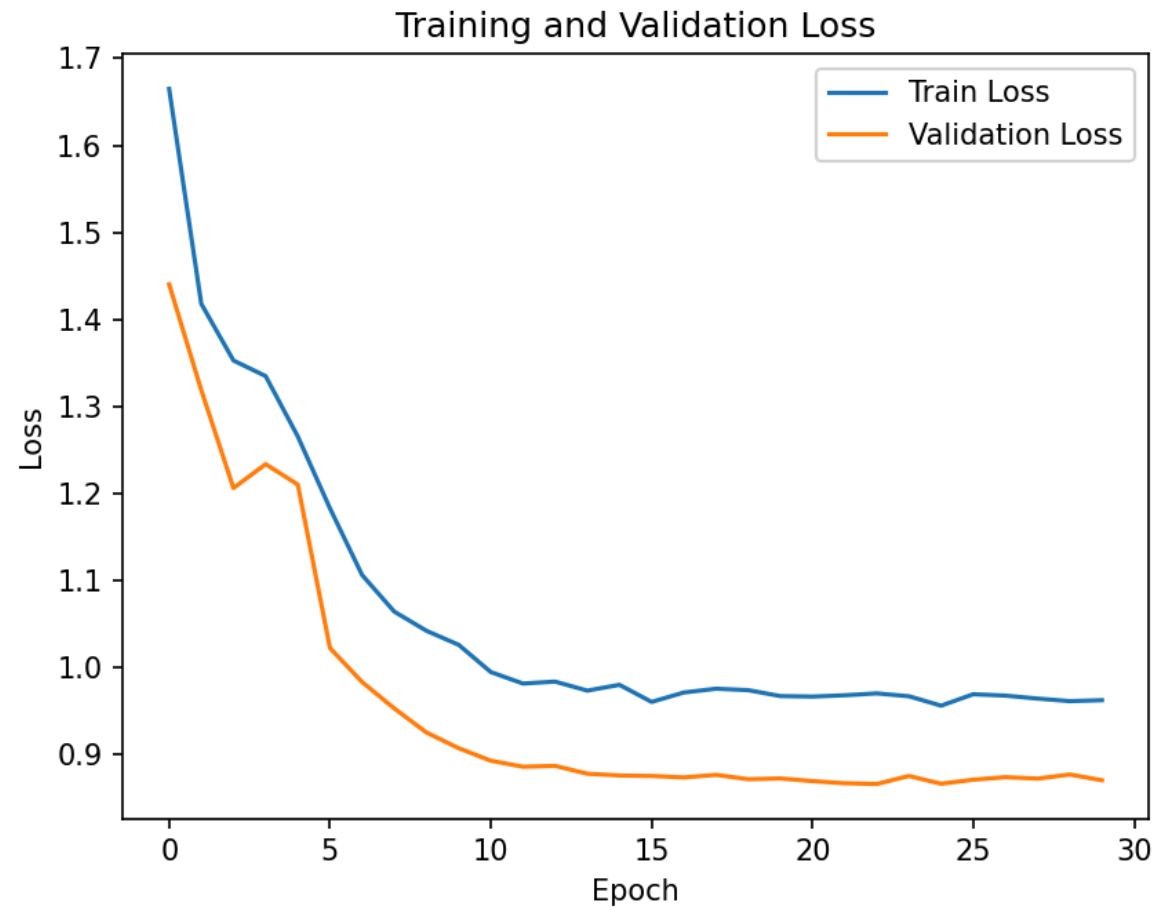- **Accuracy**: 68.39%
- **F1-Score**: 0.6780
- **Confusion Matrix**:

CNN Confusion Matrix

- **Training and Validation Accuracy:**

Training and Validation Accuracy for CNN

- **Training and Validation Loss:**

Training and Validation Loss for CNN

Convolutional Neural Network Analysis

After implementing and modifying new data preprocessing techniques and retraining the model, the Convolutional Neural Network model shows significant improvements with accuracy of {68.39}% and a F1-Score of {0.678}. The confusion matrix shows that the updated model provides some better predictions for happy, neutral, sad, and surprise classes. Misclassifications in challenging classes like fear, sad, and angry have been reduced. The training and validation accuracy curve indicates that the model is improving steadily until 63-68% where the curve flattens out. Similarly, the loss curve shows consistent decrease in both training and validation loss, indicating that the model is learning effectively on the training and validation data.

Overall Analysis

Of the three models, the CNN model demonstrates the strongest performance, obtaining the highest accuracy and F1-score, while the Random Forest and SVM models show weaker results. A key issue across all models is the significant misclassification of certain emotions. This may be due to that fact that emotions such as fear, disgust, and angry share similar features. Another factor that may cause the misclassification is the imbalance in classes. Some emotions such as happy has a larger dataset than disgust. This imbalance may lead to bias towards the majority class, making it harder for the models to accurately predict minority classes. Additionally, the 48x48 pixel size of the images may have limited the model's ability to capture fine details, constraining the amount of available visual information that could differentiate similar emotions.

Tradeoffs, Strengths, and Limitations.

- **Random Forest**: Simple and interpretable but limited by poor generalization and bias toward the majority class.
- **SVM**: Better than Random Forest but computationally intensive and struggles with imbalanced data.
- **CNN**: Best performance due to its ability to learn complex patterns, but computationally expensive and requires more data preprocessing

Next Steps

To further improve accuracy, we will transition to using the CK+ dataset. Additionally, we will use facial and hand key points from MediaPipe and map them onto a VRM model, to enable a real time transfer of facial expressions and hand movements to the virtual avatar. We also plan to introduce event-based features such as special effects, subtitiles, or pre-loaded animations.
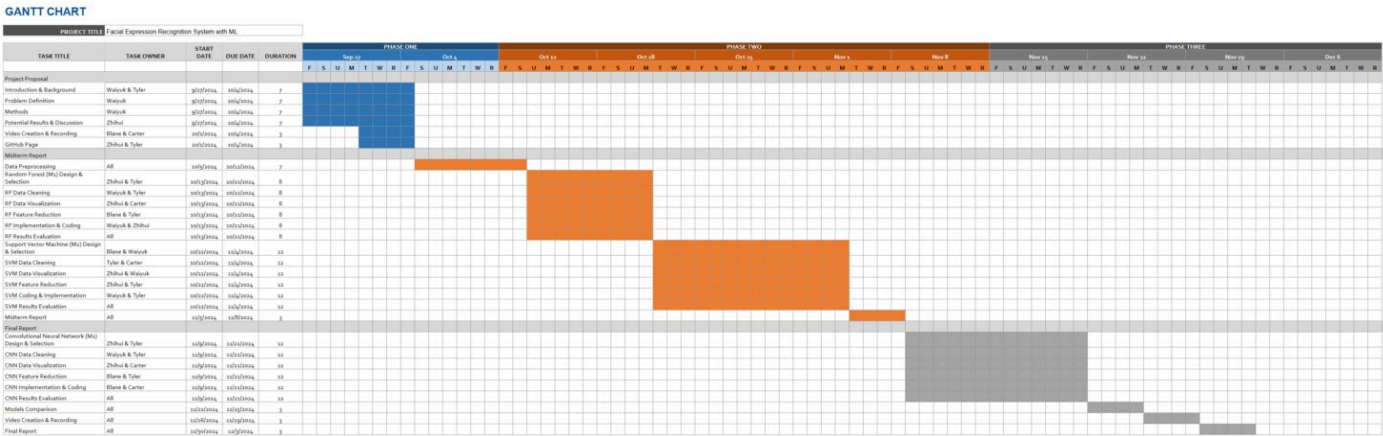
References

References

1. A. Huang and M. J. C. Samonte, "Using Emotion Analysis, Eye tracking, and Head Movement to Monitor Student Engagement among ESL Students with Facial Recognition Algorithm (Mediapipe)," 2024 7th International Conference on Advanced Algorithms and Control Engineering (ICAACE), Shanghai, China, 2024, pp. 509-513, doi: 10.1109/ICAACE61206.2024.10548871.

2. F. M. Sadikoğlu and M. Idle Mohamed, "Facial Expression Recognition Using CNN," 2022 International Conference on Artificial Intelligence in Everything (AIE), Lefkosa, Cyprus, 2022, pp. 95-99, doi: 10.1109/AIE57029.2022.00025.

3. K. Wang, R. Li, and L. Zhao, "Real-time facial expressions recognition system for service robot based-on ASM and SVMs," 2010 8th World Congress on Intelligent Control and Automation, Jinan, 2010, pp. 6637-6641, doi: 10.1109/WCICA.2010.5554164.

Contribution Table

Name	Proposal Contributions
Waikyuk Kwong	Improved CNN Model performance, Real-time detection logic, Presentation
Zhihui Chen	Conducted model comparison evaluations, Streamlit Implementation & Debugging
Tyler Lin	Streamlit Implementation & Debugging
Blane R. York	Prepared slides, Organized files, Updated file descriptions
Carter D. Robinson	Developed Capture Window, Developed Image Upload

Gantt Chart

Gantt Chart

[Gantt Chart](#)