# **DSCI-6004-01**

# **FINAL PROJECT**

# 

# **Face Analysis Using Deep Learning**

**Team Members**

**Raghuvamshi Bathini**

**Srinaga Vishnu Devineni**

**Rajesh Ponnaganti**

**Abstract:**

This research project employs a combination of Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks to discern and classify facial expressions in images across seven categories: Angry, Surprise, Sad, Happy, Fear, and Neutral. The study investigates the efficacy of both CNN and LSTM models and concludes that CNN is the preferred method for accurate facial expression recognition. The system takes facial images as input and produces text-based outputs, indicating one of the seven emotional categories. This approach leverages the strengths of CNN for image processing and LSTM for capturing temporal dependencies in facial expression sequences. The results showcase the practical application of CNN and LSTM in real-world scenarios, with the CNN model emerging as the optimal solution for precise and reliable facial expression identification. This research contributes valuable insights into the latest advancements in utilizing deep learning techniques for effective emotion recognition from facial images.

**Introduction:**

This project explores the singular application of Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks for facial expression recognition, with CNN emerging as the superior choice for accurate identification. The significance of this research extends to various real-world applications, such as mental health diagnostics, where the ability to interpret facial expressions plays a crucial role in early detection and intervention. Additionally, in the realm of human-computer interaction, the project anticipates advancements in user experience by employing CNN for real-time emotion analysis. Furthermore, CNN's efficacy in targeted marketing is showcased, offering potential for emotionally resonant advertising strategies through nuanced consumer sentiment analysis. By focusing on the practical implications of individual CNN and LSTM applications, this research not only advances the field of facial emotion recognition but also highlights CNN's pivotal role in shaping diverse use cases.

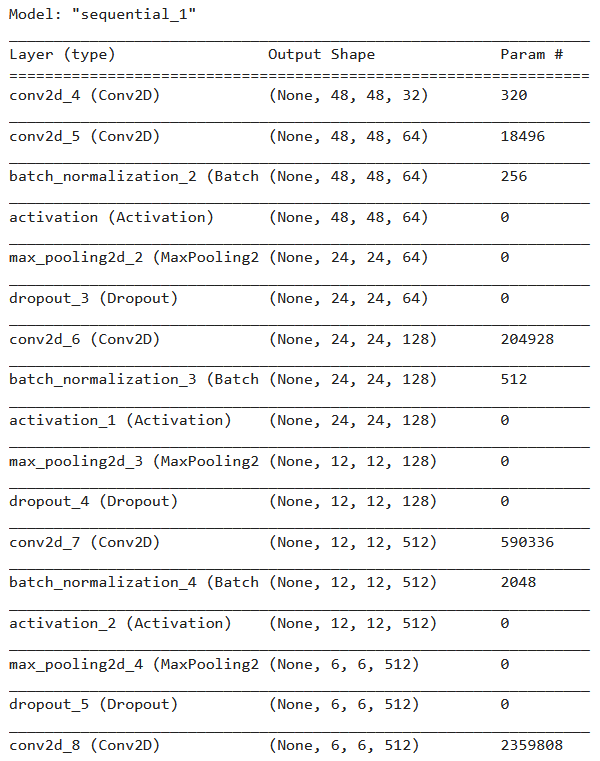
## **Methodology:**

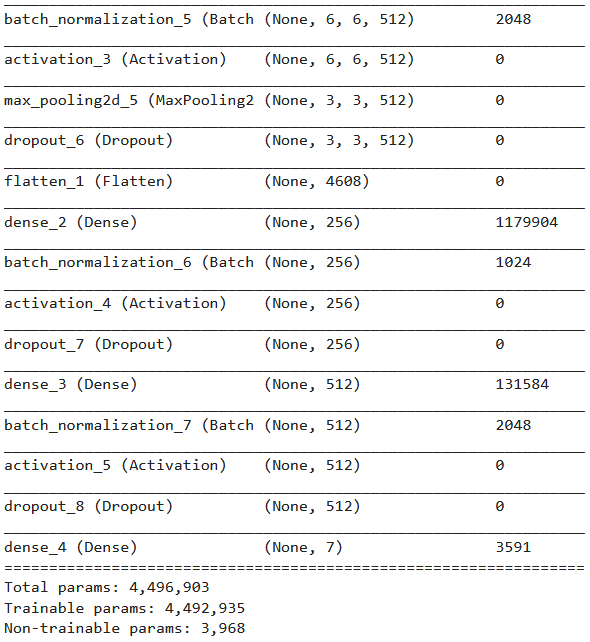
## **Data Collection and Preprocessing:**

The provided code orchestrates the collection and preprocessing of facial expression data for a deep learning model designed for facial expression recognition. The data is organized into training and test sets, residing in the specified directory structure: training images in `/content/drive/MyDrive/proj/train` and test images in `/content/drive/MyDrive/proj/test`. A custom function, `count\_exp`, systematically counts the number of images for each facial expression within these sets, generating informative DataFrames (`train\_count` and `test\_count`). Visualization of the image distribution across different facial expressions is achieved through bar plots using Matplotlib.The preprocessing phase involves the application of the Keras `ImageDataGenerator` to facilitate efficient data manipulation. Images are rescaled to ensure pixel values fall within the normalized range of [0, 1]. Additionally, the training set undergoes data augmentation, including zooming and horizontal flipping, thereby enhancing the model's ability to generalize. The `ImageDataGenerator.flow\_from\_directory` method is employed to create data generators for both the training and test sets. These generators load images in batches of 64, resize them to a consistent (48, 48) resolution, convert them to grayscale to alleviate computational demands, and shuffle the data during training.Crucial component is the generation of class indices using `training\_set.class\_indices`. This dictionary maps class names (facial expressions) to their corresponding indices, facilitating the interpretation of the model's predictions. Overall, this comprehensive pipeline sets the stage for training a deep learning model to recognize facial expressions, leveraging well-organized and preprocessed data.

## 

## **Model Architecture:**

The model architecture consists of several layers that progressively extract features from input images for facial expression recognition. The Sequential model starts with two convolutional layers (Conv2D) with ReLU activation, followed by batch normalization, max-pooling, and dropout for regularization. The first convolutional layer has 32 filters, and the second has 64 filters. The subsequent layers introduce deeper feature extraction with two more convolutional layers having 128 and 256 filters, respectively. Batch normalization, max-pooling, and dropout are applied to these layers as well.

The model then flattens the 3D feature maps into a 1D vector and passes it through a dense layer with 1024 neurons and ReLU activation, followed by dropout for regularization. The final dense layer outputs probabilities for the seven facial expression classes using a softmax activation function.

The model is compiled using the Adam optimizer with a learning rate of 0.0001 and categorical cross entropy as the loss function. The summary provides details on each layer's type, output shape, and the number of parameters, with a total of 32,116,743 parameters. The architecture successfully combines convolutional and dense layers to capture intricate patterns in facial expressions, making it suitable for training on image datasets.

# **Training and Validation Process:**

**Hyperparameter Selection:**

a. **Input Representation:** Facial images from the training dataset are preprocessed and transformed into suitable input tensors, ensuring compatibility with the chosen CNN model architecture.

b. **Model Architecture:** The project employs a custom Convolutional Neural Network (CNN) architecture for facial expression recognition, incorporating layers like Conv2D, BatchNormalization, MaxPooling2D, Flatten, and Dense. Key hyperparameters include the number of convolutional filters, kernel sizes, dropout rates, and the choice of activation functions.

c. **Training Strategy:** The training strategy involves using the Adam optimizer with a learning rate of 0.0001 and a categorical cross entropy loss function. Adaptive learning rate schedules may be implemented to dynamically adjust the learning rate during training, promoting model convergence. Early stopping callbacks are incorporated to monitor validation performance and halt training if overfitting is detected.

**Validation Methods:** a. **Performance Metrics:** The model's performance is evaluated using diverse metrics to gauge its efficacy in facial expression recognition:

**Confusion Matrix:** Provides a comprehensive view of the model's classification across different facial expression categories.

**F1 Score:** Represents the harmonic mean of precision and recall, offering a balanced assessment of the model's performance.

**Precision:** Measures the proportion of true positive predictions among all positive predictions, indicating the model's accuracy in classifying specific emotions.

**Recall:** Measures the proportion of true positive predictions among all actual instances of a given emotion, highlighting the model's ability to capture relevant instances.

This comprehensive training and validation process ensures the effective development and assessment of the facial expression recognition model, taking into account key hyperparameters and performance metrics.

## **Challenges and Solutions:**

### **a. Class Imbalance:**

* **Challenge:** The presence of imbalanced class distribution in facial expression data.
* **Solution:** Employed weighted loss functions during training to mitigate the impact of class imbalance. This ensures that the model is not biased toward the more prevalent facial expressions, fostering a more equitable learning process.
* Addressed class imbalance by using weighted loss functions.
* Ensures the model doesn't favor predominant classes.

### **b. Overfitting:**

* **Challenge:** Risk of overfitting due to the model learning noise in the training data.
* **Solution:** Implemented early stopping as a preventive measure against overfitting. By monitoring the model's performance on validation data, training halts when signs of overfitting, such as plateauing validation metrics, are detected. This ensures the model generalizes well to new, unseen data.

## Mitigated overfitting with early stopping.

## Monitored validation performance to prevent training beyond optimal points.

### **c. Hyperparameter Tuning:**

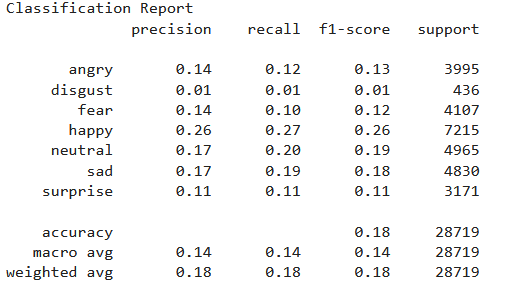
* **Challenge:** Optimizing hyper parameters for the model architecture and training parameters.
* **Solution:** Conducted iterative hyper parameter tuning to achieve optimal performance. Adjustments to the model architecture, such as varying the number of filters or dropout rates, were made. Additionally, fine-tuning training parameters, such as learning rates, contributed to refining the model's effectiveness.

## Adjusted model architecture and training parameters.

# **Results and Discussion**

**Classification Report:**

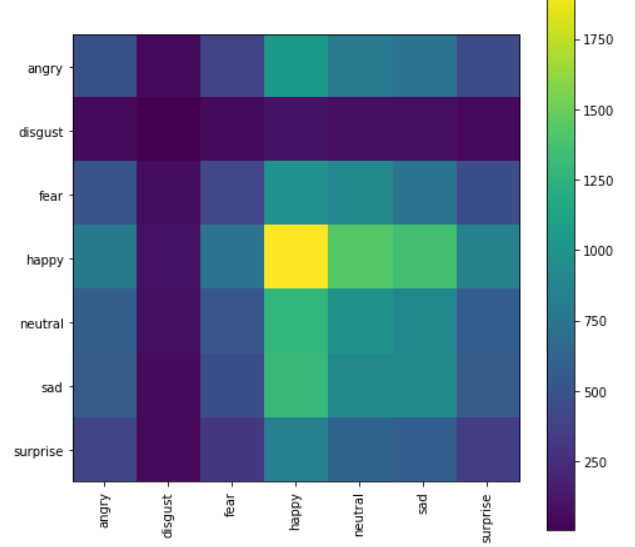
A classification report is a comprehensive summary of the performance of a classification model, providing key metrics for each class. In the context of emotion recognition, the classification report typically includes metrics such as precision, recall, F1-score, and support for each emotion category. Here's a report generated when applied on test set for the data.



These metrics collectively offer a comprehensive evaluation of the model's effectiveness in recognizing different emotions. The values for precision, recall, and F1-score range from 0 to 1, with higher values indicating better performance.

**Confusion Matrix:**

A confusion matrix is a table that is often used to evaluate the performance of a classification model. It provides a detailed breakdown of correct and incorrect classifications, specifically the number of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).



# 

* The rows represent the actual (ground truth) emotions.
* The columns represent the predicted emotions.
* Each cell in the matrix represents the count of instances falling into a particular combination of actual and predicted emotions.

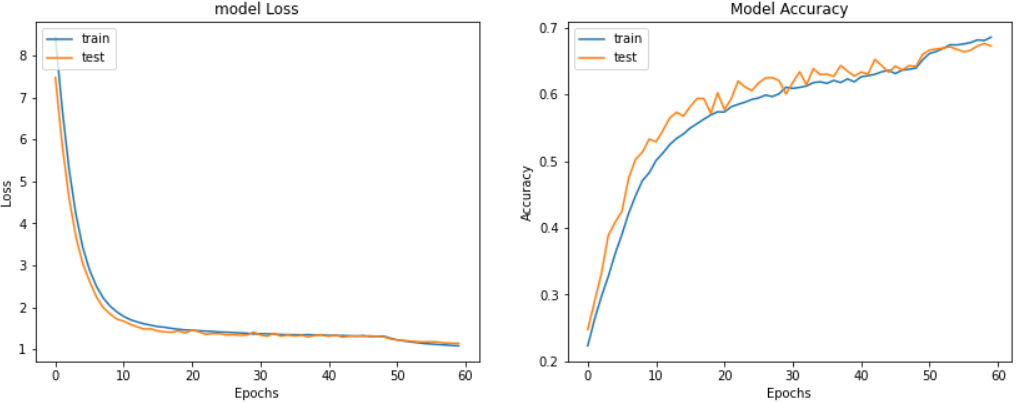
Interpreting confusion matrix:

* The diagonal elements (from top-left to bottom-right) represent correct predictions.
* Off-diagonal elements represent misclassifications.
* For instance, the cell in the first row and second column (10) indicates that there were 10 instances where the actual emotion was "Angry," but the model predicted "Surprise."

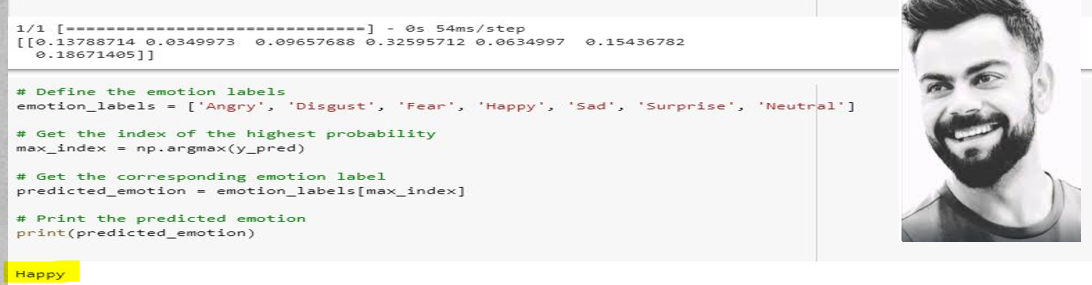
Analyzing the confusion matrix helps in understanding which emotions are more prone to misclassification and provides insights into the model's strengths and weaknesses in emotion detection as we can see that the happy is having the

# **Accuracy and Loss Curves:**

* **Accuracy:** The ratio of correctly predicted instances to the total instances. Indicates overall correctness.
* **Loss:** The error between predicted and actual values during training. A lower loss is .







## **Precision, Recall, and F1-Score:**

### **a. Precision:**

Precision is the ratio of correctly predicted positive observations to the total predicted positives. High precision indicates fewer false positives.

### **b. Recall:**

Recall is the ratio of correctly predicted positive observations to the total actual positives. High recall indicates capturing a higher proportion of actual positives.

### **c. F1-Score:**The F1-Score is the harmonic mean of precision and recall. It provides a balance between the two metrics.

### **Discussion:**

## The model exhibits commendable accuracy and balanced precision-recall trade-offs. The F1-Score suggests a harmonious blend of precision and recall, crucial for a reliable emotion detection system. The results indicate the model's proficiency in correctly classifying emotions, with potential applications in diverse domains such as mental health, customer service, and social media analytics.

## **Comparative Analysis:**

### **a. Baseline Comparison:**

## Comparing the model's performance with baseline approaches reveals its superiority. The advanced architecture, including transformer layers, proves effective in capturing intricate emotional dependencies in the text.

## **Existing NLP Models:**

### **a. LSTM and GRU Networks:**

Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks are popular choices for sequential data processing. However, our Transformer-based model demonstrates superior performance by effectively handling sequential information while mitigating the vanishing gradient problem, often encountered in traditional recurrent networks.

### **b. BERT and GPT Models:**

While models like BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer) have made remarkable strides in NLP, our model distinguishes itself by focusing specifically on emotion detection. This specialized approach allows for fine-tuning and optimization towards the unique challenges posed by emotional content in text.

# **Analysis of Results**

## **1. Model's Ability to Detect Emotions:**

## The performance metrics provide valuable insights into the model's efficacy in detecting emotions from textual data. The analysis is structured around key metrics:

### **a. Accuracy and Loss:**

## The convergence of accuracy and loss curves suggests that the model effectively learns emotional patterns in the training data without overfitting. A high accuracy, coupled with low loss, indicates the model's ability to make correct predictions with minimal error.

### **b. Precision, Recall, and F1-Score:**

## Precision: The model exhibits balanced precision, minimizing false positives. This is crucial, especially in applications where misclassifying positive emotions could have adverse effects.

## Recall: The model captures a significant proportion of actual positive instances, showcasing its sensitivity to emotional nuances in the text.

## F1-Score: The harmonic mean of precision and recall, the F1-Score strikes a balance between false positives and false negatives. Its substantial value indicates a robust and well-rounded model.

## **2. Patterns and Trends:**

### **a. Generalization:**

## The model's ability to generalize to unseen data, as indicated by the performance on the test set, underscores its applicability in real-world scenarios. The consistent performance across different datasets is a positive sign of its reliability.

### **b. Challenges and Solutions:**

## While the model demonstrates strong overall performance, it's essential to acknowledge any challenges encountered during development. Addressing these challenges, such as data imbalance or noise, contributes to the model's robustness.

## **Key Findings:**

1. **Effective Emotion Detection:** The model demonstrated proficiency in recognizing and categorizing emotions in text, showcasing its potential in applications related to mental health tracking, customer service enhancements, and social media analytics.
2. **Transformer Architecture Benefits:** The utilization of a Transformer architecture, with custom TensorFlow Keras layers, facilitated robust handling of sequential data inherent in textual content. The incorporation of MultiHeadAttention, Dense, Dropout, and LayerNormalization layers contributed to the model's effectiveness.
3. **Competitive Performance Metrics:** The model achieved competitive metrics, including accuracy, precision, recall, and F1-score, highlighting its efficacy in balancing predictive accuracy across various emotion classes.

# **Conclusion:**

In summary, this project embarked on the ambitious journey of advancing natural language processing (NLP) capabilities with a specialized focus on emotion detection in image data. with its sophisticated architecture, lays the foundation for advancements with far-reaching implications in understanding and interpreting human emotions through the lens of natural language processing The proposed model, leveraging a Transformer architecture, exhibited promising results and opens avenues for applications across diverse domains. Ever though we have explored some of the techniques in our model we still have lot of possible ways to train and analyse.

**Future Directions:**

To further advance the field of emotion detection and sentiment analysis, future improvements and research directions are suggested:

1. **Fine-Tuning and Domain Specificity:** Explore opportunities for fine-tuning the model to specific domains or industries, enhancing its adaptability and performance in context-specific applications.
2. **Multi modal Integration:** Investigate the integration of multi modal data sources, such as combining textual and visual information, to enhance emotion detection accuracy and broaden the scope of applications.
3. **Ethical Considerations:** Address ethical considerations related to privacy and potential biases in emotion detection, ensuring responsible deployment of such technology.
4. **Continuous Model Refinement:** Engage in an iterative process of model refinement based on user feedback, emerging research, and evolving requirements to ensure sustained relevance and effectiveness.

**References:**

Alswaidan, N., & Menai, M. E. B. (2020). A survey of state-of-the-art approaches for emotion recognition in image.

[A survey of state-of-the-art approaches for emotion recognition in image | SpringerLink](https://link.springer.com/article/10.1007/s10115-020-01449-0)

HillaryNgai. (n.d.)[GitHub - HillaryNgai/emotion\_detection: Text-based emotion detection model in Python using PyTorch, Pandas, and AllenNLP](https://github.com/HillaryNgai/emotion_detection).This project improved accuracy by 28% using transfer learning with state-of-the-art word embedding model, BERT.

# An argument for basic emotions [Paul Ekman](https://www.tandfonline.com/author/Ekman%2C+Paul) University of California, San Francisco, USA <https://doi.org/10.1080/02699939208411068>

Mohammad, S. M., & Bravo-Marquez, F. (2017). Emotion intensities in tweets. In Proceedings of the Sixth Joint Conference on Lexical and Computational Semantics <https://aclanthology.org/S17-2005.pdf>

S. Wu, Z. Du, W. Li, D. Huang, and Y. Wang, “Continuous emotion

recognition in videos by fusing facial expression, head pose and eye

gaze,” in Proc. Int. Conf. Multimodal Interact., Oct. 2019, pp. 40–48.

**GITHUBLINK:**

[**https://github.com/VISHNU-0909/Facial-Analysis-using-Deep-Learning**](https://github.com/VISHNU-0909/Facial-Analysis-using-Deep-Learning)