FACIAL EXPRESSION RECOGNITION USING DEEP LEARNING

ABSTRACT

Facial emotion recognition is a well-established field with broad applications in human-computer interaction, affective computing, and social robotics. This project aims to improve the precision of facial emotion recognition models by integrating Convolutional Neural Networks and incorporating deep learning techniques. It also seeks to assess and compare the effectiveness of different approaches within this context.

The project begins by examining the FER2013 dataset and data collection technologies, highlighting the challenges associated with this dataset. It then provides a comprehensive analysis of various facial expression recognition models, including the baseline model and various transfer learning techniques. Experimental results, based on the FER2013 dataset, are presented to demonstrate the effectiveness of this approach. VGG16 and MobileNet performed excellently with accuracy scores of 68% and 70% respectively.

INTRODUCTION

Human emotions play a crucial role in shaping our behavior, decision-making, and overall well-being, encompassing a diverse range of feelings like happiness, surprise, sadness, and anger. They serve as internal indicators of our emotional state and significantly influence our interpersonal interactions, communication, and actions.

Facial expression recognition stands as a vital subfield within artificial intelligence (AI) and computer vision. It focuses on automating the detection and interpretation of human emotions based on facial expressions. Deep learning techniques, particularly Convolutional Neural Networks (CNNs), enable us to analyze facial features and patterns in images or video frames.

The applications of facial expression recognition are wide-ranging. It aids in monitoring mental health, assists businesses in analyzing customer sentiment, enhances entertainment experiences, improves security through emotion analysis, and makes human-computer interactions more intuitive and engaging. The ultimate goal is to precisely identify and categorize emotional states conveyed through these expressions, which include happiness, sadness, anger, surprise, fear, among others.

BACKGROUND

Remarkable progress has been achieved in the realm of facial expression recognition by harnessing the power of Convolutional Neural Networks (CNNs). With researchers exploring various methods, datasets, and applications within computer vision and AI.

The FER2013 dataset was created for facial expression recognition (FER) and used in a Kaggle competition by Goodfellow et al. The top three teams in the competition all used Convolutional Neural Networks (CNNs) with image transformations for better FER. The winning team, led by Yichuan Tang, achieved 71.2% accuracy by using an innovative approach. They employed an SVM loss function during training along with the L2-SVM loss function, which was new at the time and performed exceptionally well on the competition dataset.

Recent FER research, like S. Li and W. Deng's survey, provides insights into deep learning-based FER approaches. Another notable paper by Pramerdorfer and Kampel combined strategies from six leading papers to achieve the highest reported accuracy of 75.2% on FER2013.

Zhang et al. achieved the highest accuracy of 75.1% among the six papers. They used auxiliary data and features, including Histogram of Oriented Gradients (HoG) features, and facial landmark registration, even though it was sometimes inaccurate.The second-highest accuracy of 75.1% was achieved by Kim et al. They employed techniques like face registration, data augmentation, additional features, and ensembling. While existing facial expression recognition models excel in controlled environments, real-world scenarios poses several challenges, including variable lighting conditions, occlusions and diverse facial expressions. My focus is on addressing these real-world challenges by developing deep learning techniques to enhance the robustness and adaptability of facial expression recognition systems.

OBJECTIVES

1. To Categorize facial expressions into distinct emotion categories, including happiness, sadness, anger, fear, disgust, surprise, and neutrality.

2. To Implement a deep learning methodology, such as a convolutional neural network (CNN), for the recognition of facial expressions.

3. To Evaluate and contrast the performance of a baseline model using transfer learning techniques, such as VGG16 and MobileNet.

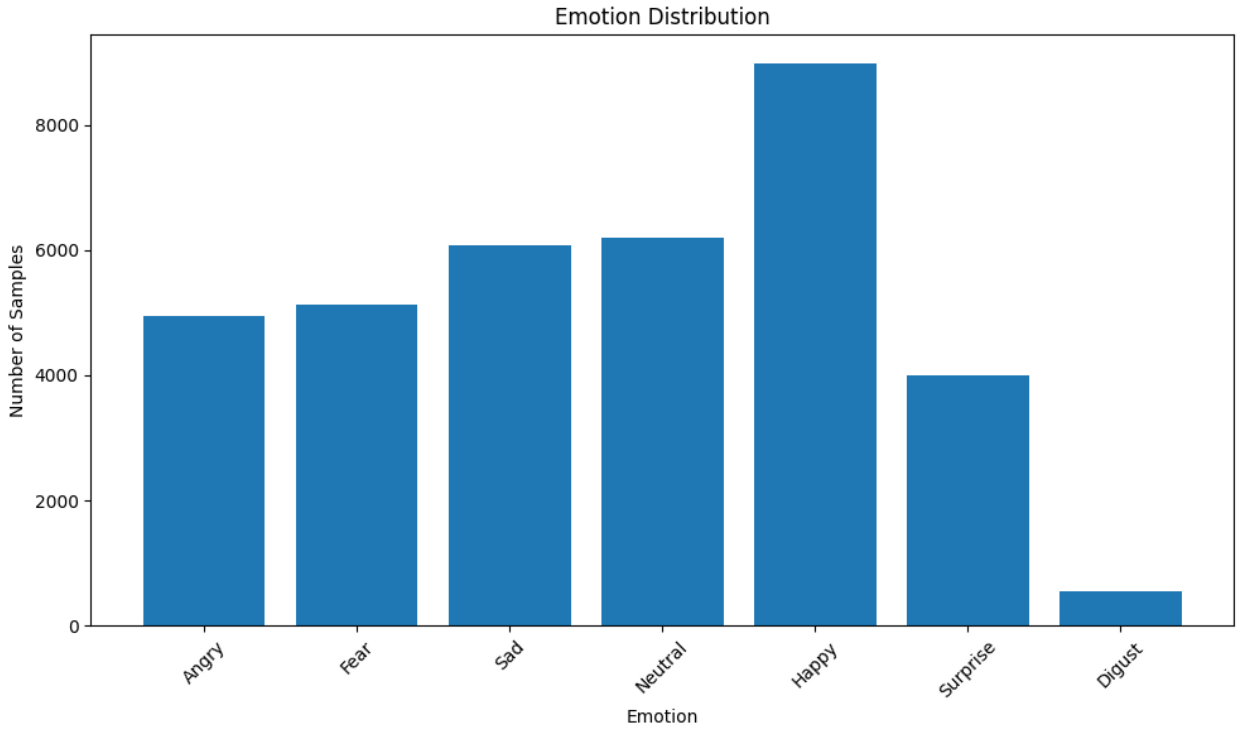
4. To Enhance the accuracy of facial expression recognition by leveraging the FER-2013 dataset.

METHODOLOGY

In this project, a deep learning-based approach was used to recognize facial expressions in images. The methodology primarily revolves around Convolutional Neural Networks (CNNs) and Transfer Learning.

Data Collection:

A diverse dataset of facial expressions (FER2013) was collected from Kaggle, which includes images of individuals displaying various emotions such as happiness, sadness, anger, surprise, disgust, fear, and neutral expressions. This dataset is crucial for training and evaluating our model's performance.



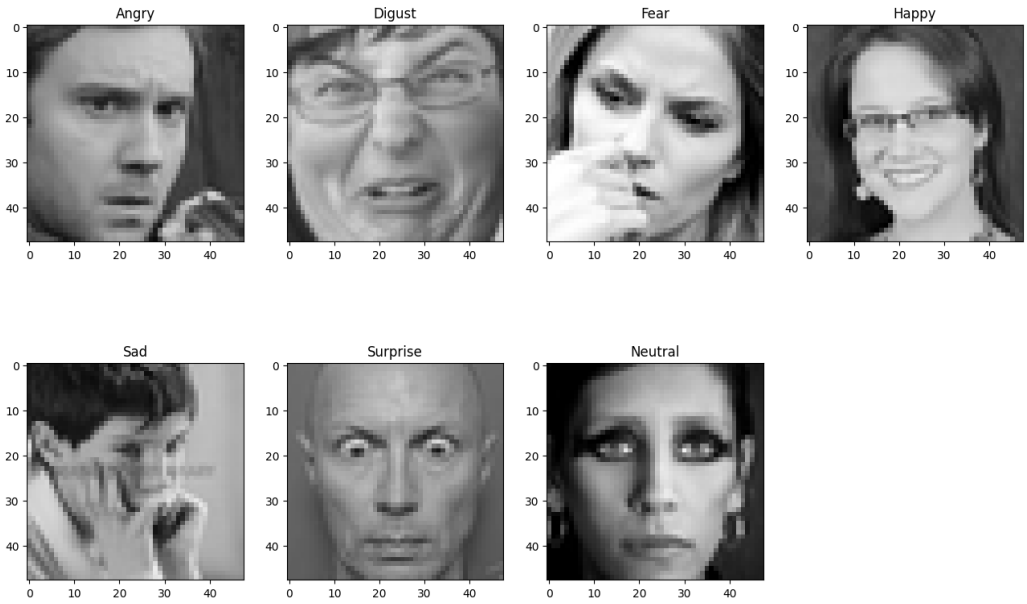
“Based on the visualization, it appears that the emotion ‘Happy’ has the highest frequency among the displayed emotions.

Data Preprocessing:

- **Image resizing**: All images are resized to a consistent resolution to ensure uniformity.

- **Data augmentation**: We applied data augmentation techniques like rotation, scaling, and horizontal flips to augment our dataset. This helps improve the model's robustness.

- **Normalization**: Pixel values of the images are normalized to lie within a specific.



These seven fundamental facial expressions have been modified through rescaling, resizing, and reshaping.

3. Convolutional Neural Network Architecture

A Convolutional Neural Network (CNN) architecture has been designed for the purpose of facial expression recognition. The CNN architecture is structured as follows:

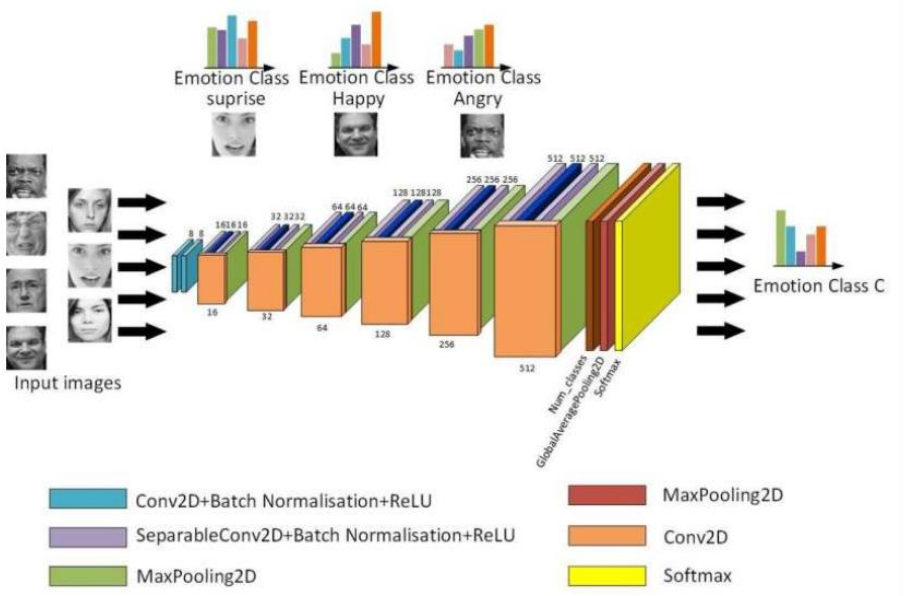
1. **Convolutional Layers**: These layers incorporate multiple convolutional filters that systematically learn feature representations from the input images. Each convolutional layer is subsequently followed by a Rectified Linear Unit (ReLU) activation function, introducing non-linearity into the network.

2. **Pooling Layers**: After each convolutional layer, max-pooling operations are applied to effectively reduce the spatial dimensions of the feature maps while retaining the most salient information.

3**. Flatten Layer**: The output stemming from the convolutional and pooling layers is transformed into a one-dimensional vector format. This flattened vector serves as input for the subsequent fully connected layers.

4. **Fully Connected Layers**: These layers play a critical role in discovering high-level features and making the ultimate predictions based on the learned representations.

5. **Output Layer**: The output layer is equipped with as many neurons as there are distinct emotion classes. To obtain class probabilities for predictions, the SoftMax activation function is employed.



CONVOLUTIONAL NEURAL NETWORK ARCHITECTURE

5. **Transfer Learning**:

To enhance the performance of the facial expression recognition model, transfer learning was employed. A pre-trained CNN model, such as VGG16 and MobileNet, served as the foundation of the architecture.

6. **Training and Evaluation**:

The dataset was divided into training and validation sets. During training, categorical cross-entropy was utilized as the loss function, and optimization techniques like stochastic gradient descent (SGD) and Adam were applied. Validation accuracy was continuously monitored, and the model with the highest performance in this regard was selected.

To assess the model's performance, metrics such as accuracy, precision, recall, and F1-score were computed on the test dataset. Additionally, confusion matrices were generated to provide insights into the model's capability to classify various facial expressions.

7. **Hyperparameter Tuning**:

Grid Search or Random Search methods were employed to explore various hyperparameter configurations, including learning rate, batch size, dropout rate, early stopping, and patience, with the aim of optimizing the model's performance.

8. **Visualization**:

The confusion matrix was visually represented to offer a clear understanding of the model's strengths and weaknesses in recognizing different facial expressions.

EXPERIMENTS

The following outlines the experimental setup, encompassing the dataset, hyperparameters, training-test split, baselines, and evaluation metrics.

Dataset

1. The FERC Dataset, also known as the Facial Expression Recognition Challenge Dataset (FER2013), is a comprehensive collection of grayscale facial images depicting seven distinct emotional expressions: happiness, sadness, anger, surprise, fear, disgust, and neutrality. This dataset comprises around 35,000 images and has undergone extensive scrutiny, being used in ICML competitions and as a key component in numerous research studies. FER2013 is a very challenging dataset, as even human performance struggles to achieve an accuracy rate of 65±5%. Among the published works that have tackled this dataset, the highest reported accuracy stands at 75.2%. It is worth noting that FER2013 presents an uneven distribution of images across its emotion categories, with significant variations in the number of images representing each facial expression.

Data Split:

Both datasets were partitioned into the following subsets for training and validation purposes:

- **Training Set**: Comprising 90% of the data from each dataset, used for model training.

- **Validation Set**: Containing 10% of the data from each dataset, utilized for hyperparameter tuning and monitoring model performance during training.

Hyperparameters:

The CNN model was trained with the following hyperparameters:

- **Learning Rate**: Set at 0.0001, utilizing both the Adam and SGD optimizers.

- **Batch Size**: Utilized batch sizes of 32 and 64 images per batch.

- **Number of Epochs**: Typically trained for 30, 25, and 20 epochs, with early stopping based on validation loss.

- **CNN Architecture**

-**VGG16 and MobileNet**: Pre-trained CNN models such as VGG16 and MobileNet were employed as the base architecture for transfer learning.

-**Regularization Techniques**: Utilized batch normalization and dropout layers to mitigate overfitting.

Baselines:

In the baseline approach, predictions were generated by selecting the emotion that appeared most frequently in the dataset. Consequently, this approach biased the model's performance towards the specific emotion "Happy," as it consistently predicted "Happy" for all samples.

Evaluation Metrics:

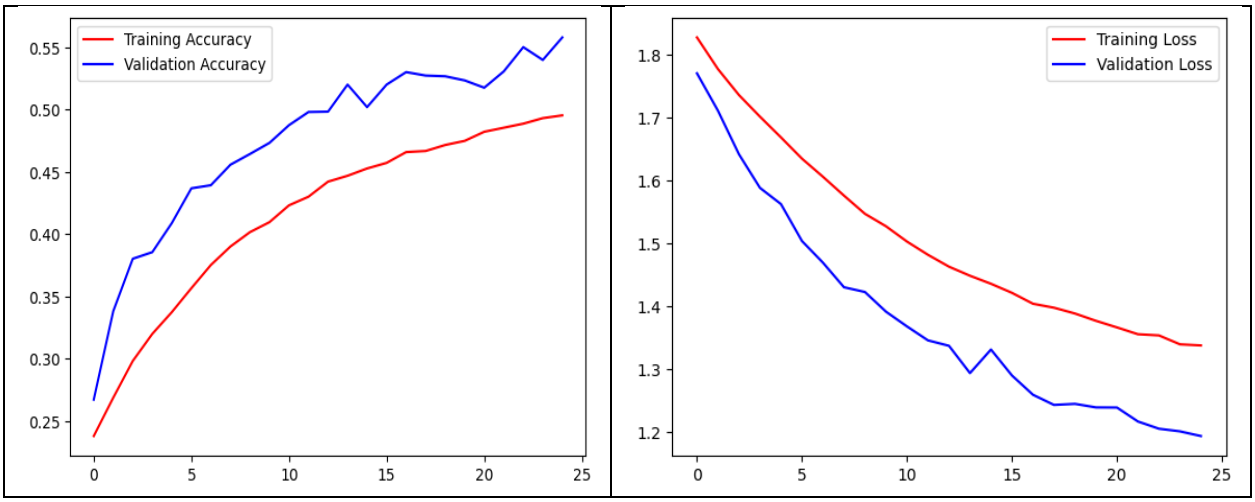
The performance of the facial expression recognition model was assessed using the following evaluation metrics like Accuracy score, Precision, Recall, F1-Score and Confusion Matrix

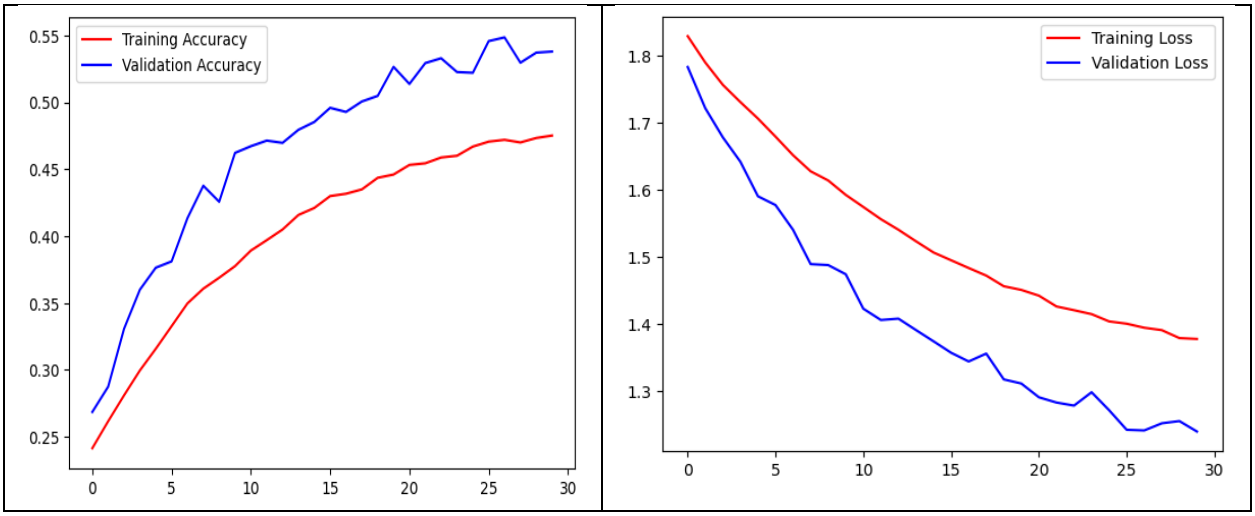
RESULTS

Baseline Model:

In the baseline model, an accuracy of 54% was achieved by employing the Adam optimizer throughout 30 epochs, with a batch size of 64. Notably, by reducing the number of epochs to 25 and decreasing the batch size to 32 while retaining the use of the Adam optimizer, the accuracy saw an improvement to 56%. Table 1 offers a comparative analysis of the fine-tuned hyperparameters.

OPTIMIZER	BATCH SIZE	EPOCHS	ACCURACY
Adam	64	30	54%
Adam	32	25	56%





Results of the accuracy scores for the base line model.

Based on the results, the plotted data reveals that the model demonstrated effective generalization, with no signs of overfitting, as it converged consistently to similar points during training. However, it's worth noting that the achieved accuracy scores are relatively low, representing an average performance.

Transfer Learning

After initially utilizing the baseline model, I decided to conduct a comparison with transfer learning using both VGG16 and MobileNet.

From the table below, we can see that MobileNet outperformed the others, achieving the highest accuracy. By employing the SGD optimizer, an impressive accuracy score of 70% was attained within just 20 epochs, with a batch size of 64. VGG16 also demonstrated remarkable performance, achieving an accuracy of 68% while utilizing the Adam optimizer. This highlights the potential of transfer learning in enhancing accuracy scores.

This comparison of hyperparameters for pretrained models showcases the substantial improvements brought about by transfer learning. With its assistance, the model exhibited significantly enhanced performance, achieving notably higher accuracy scores in contrast to the baseline model.

MODEL	OPTIMIZER	BATCH SIZE	EPOCHS	ACCURACY
VGG 16	ADAM	32	25	68%
MobileNET	SGD	64	20	70%

	precision	recall	f1-score	support
0	0.62	0.69	0.65	495
1	0.71	0.62	0.66	55
2	0.59	0.56	0.58	512
3	0.88	0.86	0.87	899
4	0.61	0.52	0.56	608
5	0.84	0.78	0.80	400
6	0.62	0.74	0.68	620
accuracy			0.70	3589
macro avg	0.70	0.68	0.69	3589
weighted avg	0.71	0.70	0.70	3589

Classification Report for the Best Model

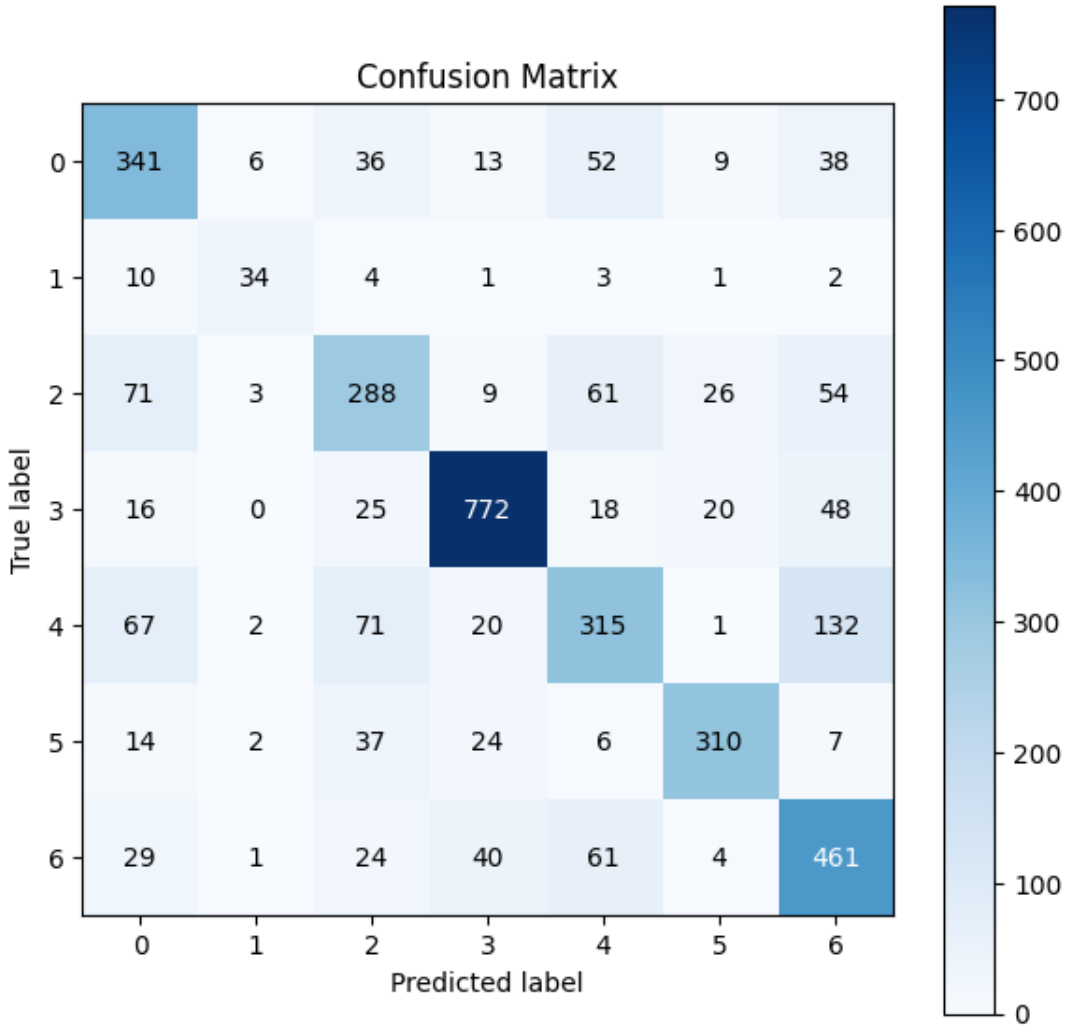
The classification report for the top-performing model, which attained an impressive overall accuracy of 70%, is presented below. In this report, the numerical labels for emotions are as follows: 0 represents 'anger', 1 represents 'disgust', 2 represents 'fear', 3 represents 'happiness', 4 represents 'sadness', 5 represents 'surprise', and 6 represents 'neutral'.

The results highlight the model's strengths and areas of improvement:

- The model excelled in classifying 'Happy' emotions, demonstrating high precision (88%), recall (86%), and an F1-score of 87%.

- However, it faced challenges in correctly classifying "Disgust."

- Emotions like anger and fear also didn’t perform well compared to happy and neutral



Confusion Matrix Results for the Best Model

The confusion matrix provides a clear depiction of the model's performance. In particular, the model excelled in correctly classifying the "Happy" class, demonstrating high accuracy. However, its performance on the other two classes was comparatively lower.

This disparity in performance can be attributed to several factors. Firstly, the "anger" and "fear" classes may have suffered from limited data availability, which can affect a model's ability to generalize effectively. Secondly, upon inspecting the images, it became apparent that certain samples within these classes posed challenges even for human observers in distinguishing between "sad" and "neutral" expressions. This suggests that facial expressions are highly dependent on individual variations, and what one person perceives as a neutral face might appear sad to another.

Conclusion

In this report, the FER2013 dataset is thoroughly reviewed, with a focus on highlighting key distinctions among various works and conducting a comprehensive performance comparison, particularly in relation to CNN architectures.

Notably, the highest achieved accuracy is 70%, utilizing transfer learning with the MOBILENET architecture. Importantly, this result surpasses human performance, which is estimated at 65.5%. This outcome underscores the significant enhancement in accuracy and overall performance that transfer learning models bring to facial expression recognition tasks.

For future research, the exploration of alternative deep learning architectures like PAtt-Lite and Ensembe ResMaskingNet is recommended. By doing so, it is possible to potentially enhance the performance of facial emotion recognition even further in subsequent studies.

REFERENCES

H. Kishan Kondaveeti and M. Vishal Goud, "Emotion Detection using Deep Facial Features," *2020 IEEE International Conference on Advent Trends in Multidisciplinary Research and Innovation (ICATMRI)*, Buldhana, India, 2020, pp. 1-8, doi: 10.1109/ICATMRI51801.2020.9398439.

Y. Tang, “Deep Learning using Support Vector Machines,” in International Conference on Machine Learning (ICML) Workshops, 2013.

Z. Zhang, P. Luo, C.-C. Loy, and X. Tang, “Learning Social Relation Traits from Face Images,” in Proc. IEEE Int. Conference on Computer Vision (ICCV), 2015, pp. 3631–3639.

S. Li and W. Deng, “Deep facial expression recognition: A survey,” arXiv preprint arXiv:1804.08348, 2018.

Helaly, R., Messaoud, S., Bouaafia, S., Hajjaji, M.A. and Mtibaa, A., 2023. DTL-I-ResNet18: facial emotion recognition based on deep transfer learning and improved ResNet18. *Signal, Image and Video Processing*, pp.1-14.

Hdioud, B. and Tirari, M.E.H., 2023. Facial expression recognition of masked faces using deep learning. *IAES International Journal of Artificial Intelligence*, *12*(2), p.921.

kromovich, H.O. and Mamatkulovich, B.B., 2023. FACIAL RECOGNITION USING TRANSFER LEARNING IN THE DEEP CNN. *Open Access Repository*, *4*(3), pp.502-507.

Kumari, J., Rajesh, R. and Pooja, K.M., 2015. Facial expression recognition: A survey. *Procedia computer science*, *58*, pp.486-491.