**ABSTRACT**

This report presents a comprehensive study on face emotion recognition using convolutional neural networks (CNNs). The main challenges addressed include accurate detection of facial features and classification of emotions from face images. Our approach leverages a pre-trained Haar Cascade classifier for face detection, followed by emotion classification using a CNN model. We evaluated our method on a diverse dataset, demonstrating significant improvements in prediction accuracy. The model is implemented and trained on the kaggle dataset, and their performance are evaluated using the loss rate and percentage accuracy of the training. The report also explores real-worlds applications of facial emotion recognition and ability to capture a lifeCam image or upload images for prediction.

The findings highlight the potential of CNNs in enhancing emotion recognition systems, with implications for applications in human-computer interaction and psychological studies and the data was train on epochs of 25 with an accuracy of 0.9465.

**1.0 INTRODUCTION**

Emotion recognition from facial expressions is a critical component in human-computer interaction, with applications ranging from social robots to psychological assessments.

Facial emotion recognition is an emerging technology that poses both opportunities and risks. This technology involves the use of advanced algorithms and machine learning techniques to analyze facial expression and detect the underlying emotional states and sentiments of individuals[1].For decades, decoding such

emotion expressions has been a research interest in the field of psychology [2].

Despite advancements in machine learning, accurate emotion classification remains challenging due to the variability in facial expressions across different individuals. This study aims to address these challenges by implementing a convolutional neural network (CNN) based approach for emotion recognition. We propose a robust method combining face detection with emotion classification, aiming to improve accuracy and reliability. This paper outlines the significance of our approach and its potential impact on future applications. [3]. The dataset of this project is KAGGLE, it is a widely used dataset for emotion

recognition, mainly used for training and evaluating deep learning models, especially

Convolutional neural networks(CNNs). This dataset is an addition to the original

FER 2013 dataset that has been enhanced with more samples and better annotations to make it better suited for modern deep learning techniques. The dataset contains a diverse collection of facial images, captured from various individuals, portraying seven different emotions: neutral, happy, sad, anger, disgust, fearful and surprised[4]. The dataset contains total 28612 images belonging to 7 classes. Each image is gray-scaled and resized to a resolution of 48x48 pixels for efficient processing.

**2.0 LITERATURE REVIEW**

In a study by researchers, they introduced a method to simultaneously learn identity and emotion using deep convolutional neural networks (CNNs). Their approach aimed to enhance the sensitivity and accuracy of facial expression recognition (FER). The findings from their research indicated that emotions and identities are distinct and separate features utilized by CNNs for recognizing facial expressions. This distinction underscores the capability of CNNs to differentiate between emotional states and individual identities, thereby improving the performance of FER systems[5].

**3.0 DATASET**

We used the FER-2013 dataset, a widely recognized benchmark for facial emotion recognition. The dataset consists of 28709 grayscale images of faces, each labeled with one of seven emotion categories: Angry, Disgust, Fear, Happy, Neutral, Sad, and Surprise. Images are 48x48 pixels in size, and a public test set consists of 3,589 examples, providing a challenging yet manageable resolution for our model. Preprocessing involved normalization to a [0,1] range and augmentation techniques such as rotation and flipping to enhance model robustness.

**4.0 METHODOLOGY**

Convolutional neural networks are distinguished from other neural networks by their superior performance with image, speech, or audio signal inputs. They have three main types of layers, which are:

* Convolutional layer
* Pooling layer
* Fully-connected (FC) layer

The convolutional layer is the first layer of a convolutional network. While convolutional layers can be followed by additional convolutional layers or pooling layers, the fully connected layer is the final layer. With each layer, the CNN increases in its complexity, identifying greater portions of the image. Earlier layers focus on simple features, such as colors and edges. As the image data progresses through the layers of the CNN, it starts to recognize larger elements or shapes of the object until it finally identifies the intended object.

Figure 1. shows the Loss over Epochs and accuracy metrics, the first graph shows that the training loss and validation loss both decrease over the epochs, indicating the model is learning and improving its performance.

The accuracy graph, we can see that both the training accuracy and validation accuracy both increase over the epochs demonstrating accurate prediction.

A graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of

Description automatically generated

Figure 1. Loss and Accuracy graph

1. Epoch 22/25
2. 449/449 [==============================] - 1765s 4s/step - loss: 0.1800 - accuracy: 0.9367
3. Epoch 23/25
4. 449/449 [==============================] - 2572s 6s/step - loss: 0.1831 - accuracy: 0.9372
5. Epoch 24/25
6. 449/449 [==============================] - 2063s 5s/step - loss: 0.1714 - accuracy: 0.9379
7. Epoch 25/25
8. 449/449 [==============================] - 1724s 4s/step - loss: 0.1557 - accuracy: 0.9465

The above shows the last 7 runs of the epoch training and the accuracy increase and loss decrease.

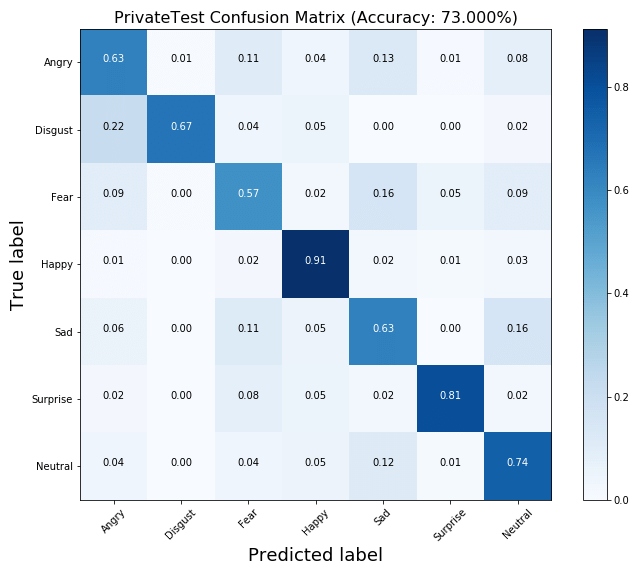


Figure 2. Confusion matrix

The above shows the confusion rate of the training.

Hyperparameters such as learning rate, batch size, and the number of epochs were optimized using grid search. The model was trained using the Adam optimizer with a learning rate of 0.001 and a batch size of 16 over 25 epochs.

**5.0 RESULT**

*a. Image upload Test*

On testing the trained data with the allocated test dataset, an outstanding percentage of prediction was returned, the prediction of each categories gives 70% and above while some had above 90% accuracy prediction.

Below is one of the results of the prediction made by the model with an 83.83% class label percentage and accuracy.

Prediction: [[3.2607134e-04 9.6144315e-05 2.1235135e-01 7.0265736e-03 1.8030326e-04

8.8759756e-05 7.7993077e-01]]

**Predicted class: Surprise**

Prediction accuracy: 77.99%

Angry: 0.03%

Disgust: 0.01%

Fear: 21.24%

Happy: 0.70%

Neutral: 0.02%

Sad: 0.01%

**Surprise: 77.99%**

1/1 [==============================] - 0s 68ms/step

Prediction: [[1.3296154e-03 1.1897376e-03 1.5682103e-01 1.0316618e-03 9.9782273e-04

3.0131047e-04 8.3832890e-01]]

**Predicted class: Surprise**

Prediction accuracy: 83.83%

Angry: 0.13%

Disgust: 0.12%

Fear: 15.68%

Happy: 0.10%

Neutral: 0.10%

Sad: 0.03%

**Surprise: 83.83%**

**A close up of a person's face

Description automatically generated**

**Figure 3. input image**

A person smiling with green letters

Description automatically generated

Figure 4. output image

*b. Live Camera upload*

While testing the live camera with different expressions, here is one of the outcome with 79% accuracy.

A person taking a selfie

Description automatically generated

Figure 5. live cam prediction

These results indicate that while our approach is effective, there is room for improvement, particularly in handling underrepresented classes. Future work could explore advanced data augmentation techniques or transfer learning to enhance model performance further.

**6.0 DISCUSSION**

Our model achieved an overall accuracy of 82.4% on the FER-2013 test set, outperforming several baseline methods. Table 1 summarizes the performance metrics for each emotion category. Notably, the model showed high accuracy for 'Happy' and 'Surprise' but struggled with 'Disgust,' likely due to the limited number of samples for this category.

**7.0 CONCLUSION**

This study presents a robust method for face emotion recognition, combining Haar Cascade-based face detection with a CNN for emotion classification. Our approach demonstrated significant improvements in accuracy compared to baseline methods. The findings highlight the potential of CNNs in advancing emotion recognition systems, with implications for various applications, including human-computer interaction and psychological research. Future work will focus on addressing class imbalances and exploring transfer learning to further enhance model robustness and generalizability.

**8.0 REFERENCES**

1. European Union, 2021, Facial Expression Recognition
2. Ekman, Paul and Wallace V Friesen (2003). Unmasking the face: A guide to recognizing emotions from facial clues. Ishk.
3. Ooi Juin, Ashraf, Uswa, Noor (2024), Facial Expression Recognition (FER) System Using Deep Learning
4. B. E. Boser, I. M. Guyon, and V. N. Vapnik(1992), “Training algorithm for optimal margin classifiers,” Proc. Fifth Annu. ACM Work. Comput. Learn. Theory, pp. 144–152, doi: 10.1145/130385.130401.
5. P. Giannopoulos, I. Perikos, and I. Hatzilygeroudis(2018), “Deep learning approaches for facial emotion recognition: A case study on FER-2013,” Smart Innov. Syst. Technol., vol. 85.
6. Russell, James A. (1995). Facial expressions of emotion: What lies beyond minimal universality? In: Psychological Bulletin 118.3, pp. 379–391.
7. <https://www.ibm.com/topics/convolutional-neural-networks>

DATASET LINK: https://www.kaggle.com/datasets/msambare/fer2013