

# FACIAL EMOTION RECOGNITION

**DHRUVKUMAR GIRISHCHANDRA PATEL**

A project report  
submitted in conformity with the  
Requirements for the degree

**Master's of Science in Information Technology**  
**Concordia University of Edmonton**  
**Faculty of Graduate Studies**  
Edmonton, Alberta  
August 2023



# FACIAL EMOTION RECOGNITION

DHRUVKUMAR PATEL

**Approved:**

Nasim Hajari, Ph.D.

---

Supervisor

August 14, 2023

---

Committee Member

August 14, 2023

Patrick Kamau, Ph.D.

---

Dean of Graduate Studies

August 14, 2023

## Abstract

Facial emotion recognition has gained substantial attention in past decade's as a key factor for human-computer interaction. Emotion recognition refers to the ability to identify and understand human emotions based on various cues, such as facial expressions, voice tone, body language, and physiological signals. This report presents a comprehensive study of facial emotion recognition using the FER2013+ dataset, providing an overview of its structure, size and different emotion categories. There are several deep learning architectures for facial emotion recognition, the method that is used for this project is Convolutional Neural Network(CNN). The model is implemented and trained on the FER2013+ dataset, and their performance are evaluated using various metrics like precision, recall, f1-score. The report also explores real-worlds applications of facial emotion recognition including emotion-aware user interfaces, investigation purposes, personalized recommendation systems, and mental health monitoring tools.

**Keywords:** Emotion recognition, CNN, Deep learning, dataset, precision, recall, f1-score.

# Contents

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Introduction</b>                                  | <b>1</b>  |
| <b>2</b> | <b>Motivation</b>                                    | <b>2</b>  |
| <b>3</b> | <b>Literature review</b>                             | <b>3</b>  |
| <b>4</b> | <b>Dataset</b>                                       | <b>4</b>  |
| <b>5</b> | <b>Project Design</b>                                | <b>5</b>  |
| <b>6</b> | <b>Evaluation</b>                                    | <b>6</b>  |
| 6.1      | Training and Validation: Accuracy and Loss . . . . . | 6         |
| 6.2      | Confusion Matrix . . . . .                           | 7         |
| 6.3      | Classification Report . . . . .                      | 8         |
| <b>7</b> | <b>Experiments</b>                                   | <b>10</b> |
| <b>8</b> | <b>Limitations or Challenges</b>                     | <b>12</b> |
| <b>9</b> | <b>Conclusions</b>                                   | <b>13</b> |

# List of Tables

|   |  |   |
|---|--|---|
| 1 | Accuracy of Previous Systems . . . . . | 3 |
| 2 | Classification Report . . . . .        | 8 |
| 3 | Individual Emotion Accuracy . . . . .  | 9 |

## List of Figures

|    |  |    |
|----|--|----|
| 1  | Data distribution . . . . .                          | 4  |
| 2  | CNN model architecture . . . . .                     | 5  |
| 3  | User Interface . . . . .                             | 6  |
| 4  | Training and Validation: Accuracy and Loss . . . . . | 7  |
| 5  | Confusion Matrix . . . . .                           | 7  |
| 6  | Emotion Results 1 . . . . .                          | 10 |
| 7  | Emotion Results 2 . . . . .                          | 10 |
| 8  | Multiple Faces with different emotions . . . . .     | 11 |
| 9  | Neutral emotion in Low lighting . . . . .            | 11 |
| 10 | Surprise emotion with mouth covering . . . . .       | 12 |

# 1 Introduction

Emotion recognition refers to the ability to identify and understand human emotions based on various cues, such as facial expressions, voice tone, body language, and physiological signals. Facial emotion recognition also known as facial expression recognition, is a significant area of research within computer vision and affective computing. Facial Emotion recognition specifically focuses on analyzing and interpreting emotions through facial expressions. It involves detecting and interpreting different facial cues, such as changes in muscle movements, eyebrow position, eye gaze, mouth shape, and other facial features. It uses the automated recognition and categorization of human emotions based on facial expressions seen in pictures or videos. In order to improve human-computer interactions and enable more emotionally intelligent applications, this technology is essential for enabling machines to understand and react to human emotions[1].

Facial emotion recognition is a fast evolving field with the center focus of computer vision, machine learning, and affective computing. Due to its numerous applications in fields like marketing, human-robot interaction, virtual reality, and healthcare, among others, emotion recognition from facial expressions has drawn a lot of interest[2].

Over last decade, advancements in deep learning algorithms and the increase in availability of large scale data-sets has increased significantly in facial emotion recognition. Convolutional neural networks(CNNs), have showed impressive capabilities in extracting distinguishing features from facial images, which results in improved accuracy of emotion classification[3].

The dataset of this project is FER 2013+, 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[4] that has been enhanced with more samples and better annotations to make it better suited for modern deep learning techniques[5]. 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[6]. The dataset contains total 65520 images belonging to 7 classes. Each image is gray-scaled and resized to a resolution of 48x48 pixels for efficient processing.

In this report, the aim is to develop a website in which users can detect emotions from their images, videos or live cameras. The project thoroughly investigate the methodologies and advancements in facial emotion recognition, with a focus on the use of the FER 2013+ dataset. Modern deep learning approaches will be examined and contrasted, along with data pre-processing methods. The ethical issues of privacy, bias, and the responsible use of emotion recognition systems will also be covered.

## 2 Motivation

Facial emotion recognition is an evolving market. Both emotion recognition and facial expression recognition have applications in various fields, including psychology, human-computer interaction, market research, entertainment, and healthcare.

- Facial emotion recognition improves human-computer interaction by adjusting responses based on users' emotional states, enabling personalized and intuitive interactions.
- IT aids in diagnosing and treating mental health conditions, monitoring emotional well-being, and early detection of mood disorders, improving mental health support.
- IT helps businesses understand customer reactions to products, advertisements, and services, enabling tailored marketing strategies and better customer engagement.
- IT aids educators in understanding student engagement and responses, enabling effective teaching methods and learning environments.
- In robotics enhances human interactions, making them more natural and intuitive in caregiving, companionship, and customer service roles.
- IT aids security systems by analyzing real-time emotional states, detecting potential threats and suspicious behavior.
- IT improves entertainment industry by enhancing video games, virtual reality, and animated characters' responsiveness to players' emotions.
- IT enhances autonomous driving by monitoring drivers' emotional states, enabling vehicle systems to respond to agitation or distraction for safety.
- This technology enhances accessibility for disabled individuals by enabling non-verbal communication through facial expressions and eye movements.

The primary objective is to create a comprehensive platform capable of accurately detecting and interpreting users' emotional states from visual data they provide. To achieve this, a sophisticated web application has been developed, leveraging advanced facial emotion recognition technology. By analyzing the intricate nuances of users' facial expressions captured through visual input, this application aims to discern a wide spectrum of emotions, providing a deeper understanding of users' feelings and reactions. This endeavor not only strives to enhance human-computer interaction by tailoring responses to users' emotional cues but also holds potential in diverse domains such as mental health assessment, personalized marketing, and immersive virtual experiences.



### 3 Literature review

According to various studies, nonverbal components convey two-thirds of human communication and verbal components one-third, with people generally inferring the emotional states of others, such as joy, sadness, and anger, using their facial expressions and vocal tones. [7], [8] In a study by [9], they proposed an approach to learn identity and emotion jointly. They used deep convolutional neural networks (CNNs) to increase the sensitivity of facial expressions and their better recognition. From their study, they concluded that emotions and identifications are different and separate features, which are being used by CNNs for Facial expression recognition (FER). They deduced a statement that expression and identity can be both used to deep learned tandem facial expression (TFE) feature and can be used to form a new model. Experimental results from this study presented the fact that this model approach achieved 84.2 percentage accuracy on FER+ Dataset. Identity and emotion combined model was experimented using different methods including ResNet18, ResNet18+FC, and TFE Joint Learning. They gave an accuracy of 83, 83 and 84 percentage respectively as seen in table 1. From different studies, different models were studied for Old FER2013, and New FER+ database models. Results of different models on old FER and FER+ are given in table below.

Table 1: Accuracy of Previous Systems

| Dataset  | Methods            | Accuracy |
|----------|--------------------|----------|
| FER 2013 | DLSVM-L2[8]        | 71       |
|          | Zhou et al.[8]     | 69       |
|          | Maxim Milakov[8]   | 68       |
|          | Radu+Marius+Cristi | 67       |
| New FER+ | This Study         | 71       |
|          | VGG13(MV)[8]       | 83       |
|          | TFE-JL[8]          | 84       |

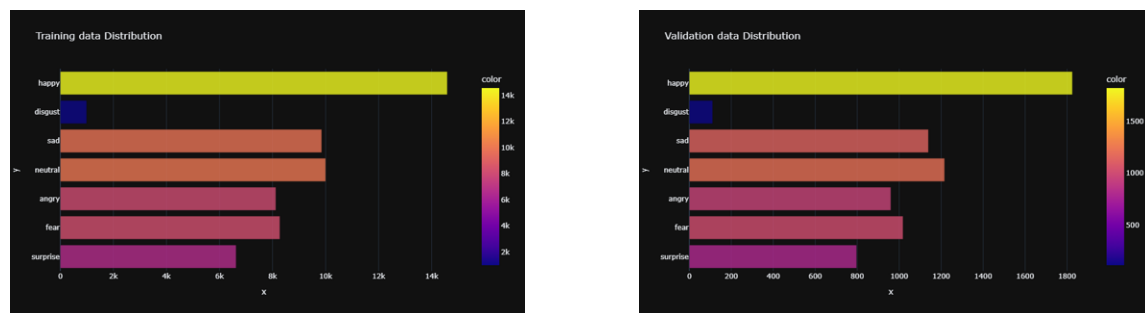
In another study [10] Based on the features of the human face, the database was able to distinguish five human emotions—happiness, anger, grief, surprise, and neutral—with an average recognition accuracy of 81.6 percentage. [11] In a another study, Eigen spaces and a dimensionality reduction technique were used to identify the fundamental emotions of sadness, anger, contempt, fear, happiness, and surprise in people’s facial expressions. [12] The system that was created had an accuracy rate for recognition of 83 percent. A different article’s research extracts local face features using principal component analysis, classifies facial expressions using an artificial neural network, and uses a unique method dubbed Canny. According to a research, the method average level of facial emotion categorization accuracy is 85.7 percent on FER+.

## 4 Dataset

In 2013 a dataset was created for facial emotion recognition named as FER-2013[4]. This dataset was small and had numerous flaws. The FER+ dataset was introduced in 2016 [13]in this paper, the authors describe the process of creating the FER+ dataset, which involved refining the emotion labels present in the FER 2013 dataset and obtaining probability distributions to capture the uncertainty associated with each label. The authors conducted a study in which human annotators adjusted the emotion labels of the images, resulting in more accurate annotations.

The project uses FER 2013+[15] as its dataset. In total 65520 images are in the dataset. Data pre-processing techniques are used for increasing the accuracy of the model[16]. Data pre-processing involves resizing the image to 48x48 pixels so that when the images from the dataset is taken as input it reduces memory usage and increases the speed of training[17]. Grayscaleing is also done on the dataset which results in images having single channel which further leads to faster training and gray-scaling also increases memory efficiency[18].

The dataset is divided into two parts training set and validation set, training set consists of 90 percentage fig. 1a of the total dataset which is 58454 images and the rest 10 percentage fig. 1b which is 7066 images are taken as validation set. The dataset includes 7 different emotion classes: Anger, Disgust, Fear, Happy, Sad, Surprise, Neutral. In fig. 1a and fig. 1b it can be seen that happiness had the highest number of images followed by neutral, sad, fear, angry, surprise and lastly disgust having least number of images. This difference in number of images as input will also result in variance in emotion accuracies.



(a) Training data distribution

(b) Validation data distribution

Figure 1: Data distribution

## 5 Project Design

Convolutional neural network(CNN) is used as the model for Facial emotion recognition in this project. CNNs has spatial hierarchical feature means it automatically detects patterns on the face provided if it within the screen[19]. CNNs provides Translation Invariance meaning emotion can be recognized irrespective of their location in the image by using shared weights in convolutional layers [20]. CNNs also provides Feature Hierarchies in which the model learns from the corners/edges first and progressively learn higher-level features in deeper layers. This helps in enhancing the models ability to discriminate emotions effectively[22].

The CNN model architecture used in this project has four convolutional layers, each followed by Batch Normalization to improve training efficiency[24]. An activation function ReLU(Rectified Linear Unit) is applied after each convolutional layer to introduce non-linearity. MaxPooling is used to downsample the spatial dimensions, reducing the computational complexity and preventing overfitting. Dropout layers are added after each MaxPooling layer to randomly deactivate some neurons during training, further preventing overfitting [21]. The output of the last MaxPooling layer is flattened into a 1D vector, which serves as input to the fully connected layers.

The model has three fully connected layers, each followed by Batch Normalization, ReLU activation, and Dropout. The fully connected layers help in capturing higher-level patterns from the extracted features [19]. The dropout layers are used to reduce overfitting during training. The final layer is a dense layer with a Softmax activation function. It has seven neurons, one for each emotion class, and it produces the probability distribution over the classes for each input image.

The following fig. 2 is the visual representation of the CNN model architecture.

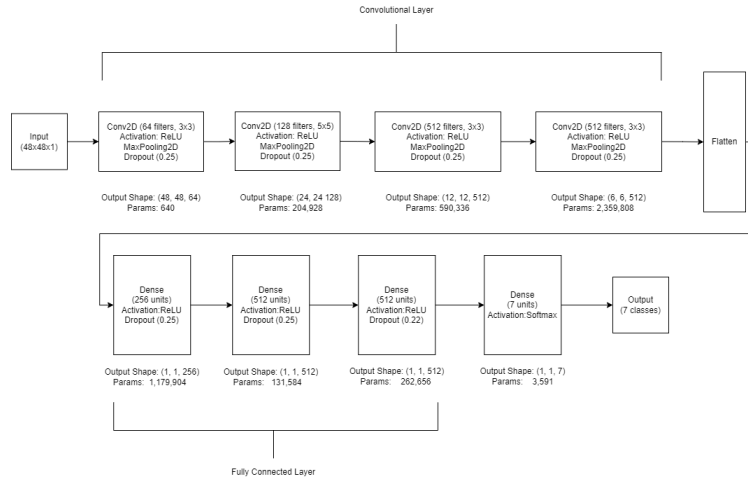
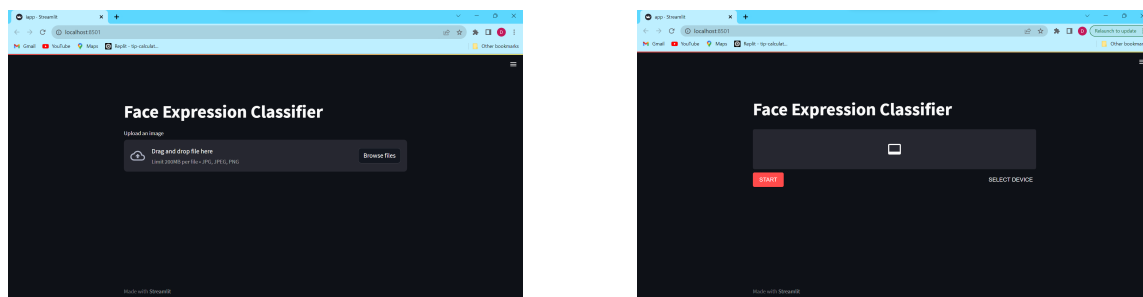


Figure 2: CNN model architecture

The final layer is a dense layer with a Softmax activation function. It has seven neurons, one for each emotion class, and it produces the probability distribution over the classes for each input image. The model uses Adam optimizer, a variant of stochastic gradient descent (SGD), which adapts the learning rate for each parameter during training [23]. The learning rate is set to 0.0001. The loss function chosen is categorical cross-entropy, which is appropriate for multi-class classification problems.

Furthermore, to display the emotion recognition system a web application using Streamlit is developed. Streamlit is a user-friendly Python library that allows us to build interactive web applications easily. Streamlit supports python libraries like OpenCV(cv2), Tensorflow and WebRTC. Below fig. 3, represents the user interface of the system through which the user can operate to detect facial emotions either from image/videos or live cameras.



(a) Interface for Images, videos

(b) Interface for live camera

Figure 3: User Interface

## 6 Evaluation

Model evaluation is an essential component of any machine learning project, and properly reporting the evaluation results is critical for effectively communicating the model’s performance. The evaluation used to assess the model’s performance are Accuracy, Precision, Recall, F1-score and Confusion Matrix. As mentioned 90 percentage dataset is used for training and the rest 10 percent is used for validation.

### 6.1 Training and Validation: Accuracy and Loss

**Training and Validation Loss:** In deep learning, training loss and validation loss are frequently used metrics. They depicts the difference between the model’s predicted output and the actual target output [25]. The left graph of fig. 4 depicts the Training and Validation loss.

**Training and Validation Accuracy:** A classification model’s performance is assessed using a performance metric called accuracy. Out of all the predictions made by the model, it calculates the percentage of accurate predictions. The model is assessed using training data to determine training accuracy and validation data to

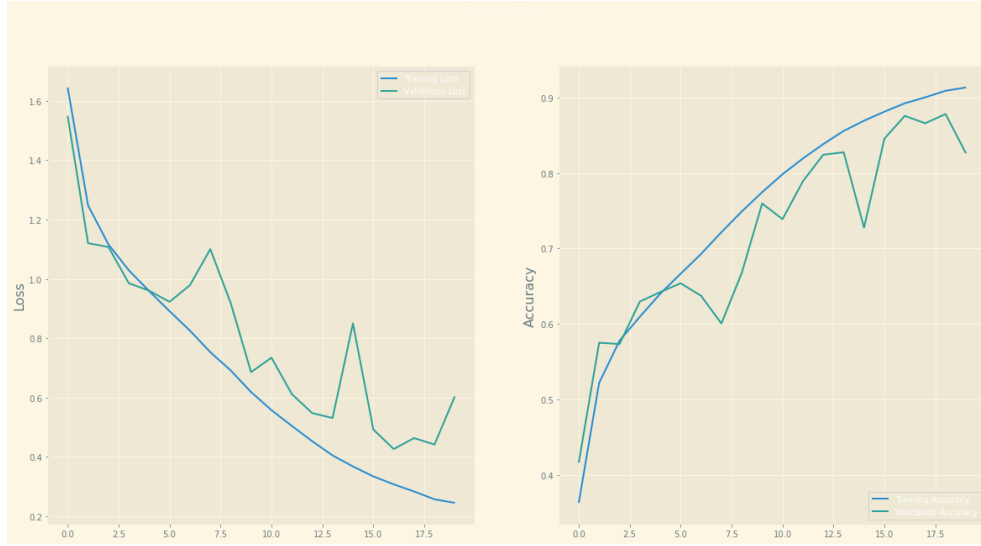


Figure 4: Training and Validation: Accuracy and Loss

determine validation accuracy during training [26]. The right graph of fig. 4 depicts the Training and Validation Accuracy.

## 6.2 Confusion Matrix

The effectiveness of a classifier on a multi-class classification task is represented by a confusion matrix fig. 5. The true classes are represented by each row in the matrix, while the predicted classes are shown by each column. The top-left to bottom-right diagonal elements of the matrix represent correctly classified samples for each class, whereas the off-diagonal elements represent incorrect classifications [27].

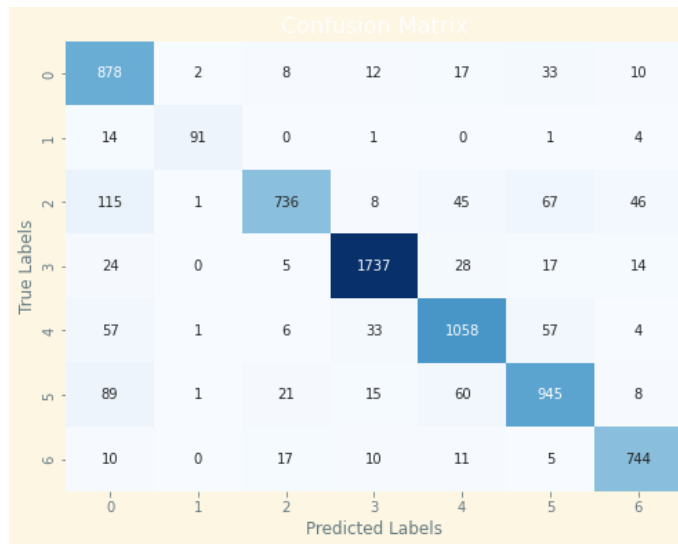


Figure 5: Confusion Matrix

The fig. 5 interprets as 878 labels were predicted correctly for class 0, while 2 labels were predicted as class 1, 8 labels as class 2, 12 labels as class 3 and so on. After all labels were predicted for each class total 6,189 labels were predicted correctly which results in 87.6 percentage accuracy.

### 6.3 Classification Report

The classification report offers a thorough summary of different evaluation metrics for each class in the dataset. After the CNN model has been trained, it is typically used to evaluate how the model performed on a validation set. For each class, the classification report includes the following crucial metrics given in the table 2

Table 2: Classification Report

| Sr no.       | Precision | Recall | F1-score | Support |
|--------------|-----------|--------|----------|---------|
| 1            | 0.74      | 0.91   | 0.82     | 960     |
| 2            | 0.95      | 0.82   | 0.88     | 111     |
| 3            | 0.93      | 0.72   | 0.81     | 1018    |
| 4            | 0.96      | 0.95   | 0.95     | 1825    |
| 5            | 0.87      | 0.87   | 0.87     | 1216    |
| 6            | 0.84      | 0.83   | 0.83     | 1139    |
| 7            | 0.90      | 0.93   | 0.91     | 797     |
| Accuracy     | -         | -      | 0.88     | 7066    |
| Macro avg    | 0.88      | 0.86   | 0.87     | 7066    |
| Weighted avg | 0.88      | 0.88   | 0.88     | 7066    |

- **Precision:** Precision measures the accuracy of positive predictions for each class [28]. For instance, for class 0 in table 2, the model’s predictions are 74 percentage accurate, meaning that 74 percentage of the predicted positive samples for class 0 are correct.
- **Recall:** Recall, also known as sensitivity, quantifies the model’s ability to correctly identify positive samples for each class [26]. For example, for class 3 in table 2, the model can recall 95 percentage of the actual positive samples in the dataset.
- **F1-score:** The F1-score is the harmonic mean of precision and recall. It provides a balanced measure of a model’s performance by considering both precision and recall. Higher F1-scores indicate better overall performance [28]. For example, class 3 in table 2 has an F1-score of 0.95, which is quite high.
- **Support:** Support represents the number of actual samples in the test set for each class [29]. It shows the distribution of samples across the different classes.
- **Accuracy:** The accuracy is the overall performance metric that measures the ratio of correctly predicted samples to the total number of samples in the test

set [26]. In this case table 2, the model achieved an accuracy of 88 percentage, indicating that it correctly predicted 88 percentage of the samples in the test set.

- **Macro Avg:** The unweighted mean of precision, recall, and F1-score for all classes is calculated by the macro average. It contributes equally to every class, regardless of size [26]. The macro-averaged precision, recall, and F1-score in this instance are all close to 86 percentage.
- **Weighted Avg:** The precision, recall, and F1-score across all classes are calculated using a weighted average, where the weights are determined by the support (number of samples) for each class. It takes into account the dataset’s class imbalance [28]. The weighted average precision, recall, and F1-score in this instance are all close to 88 percentage.

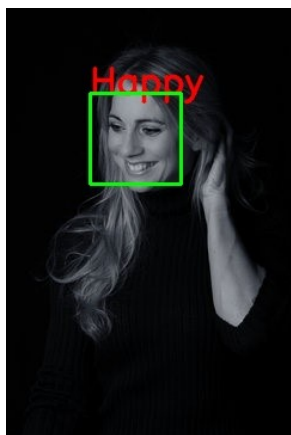
The table 3 shows the accuracy of each emotions, it is seen that happiness had highest accuracy among all other emotions 94.3 percent whereas disgust had least number of accuracy 85.6. The accuracy gap between these emotions is because happiness had higher number of images as inputs which results in better training for happy emotion and whereas disgust had the least number of image because it had least number of images as input. Other emotion had accuracies of Surprise - 92.7, Neutral - 89.8, Sad - 87, Anger - 86.4 and Fear - 86.3, Hence, the facial features of disgust emotions where not as precise.

Table 3: Individual Emotion Accuracy

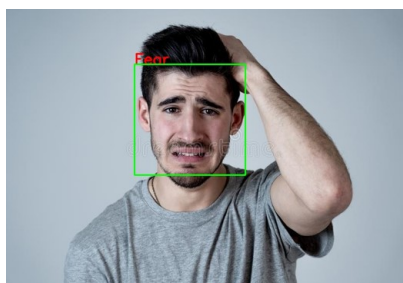
| Emotion  | Accuracy |
|----------|----------|
| Anger    | 86.4     |
| Disgust  | 85.6     |
| Fear     | 86.3     |
| Happy    | 94.3     |
| Neutral  | 89.8     |
| Sad      | 87       |
| Surprise | 92.7     |

## 7 Experiments

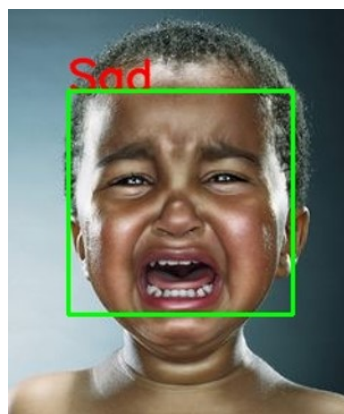
The fig. 6 and fig. 7 shows the results of the system. These images are taken from within the dataset. Hence, the accuracy of the emotions is higher. All the emotions like Happy, Fear, Sad, Neutral, Surprise were correctly detected.



(a) Happy Emotion

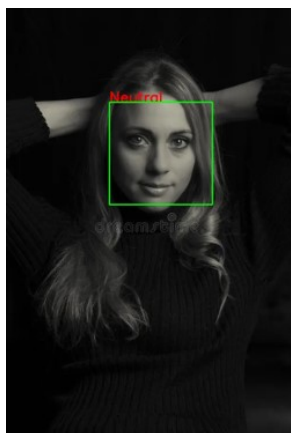


(b) Fear Emotion

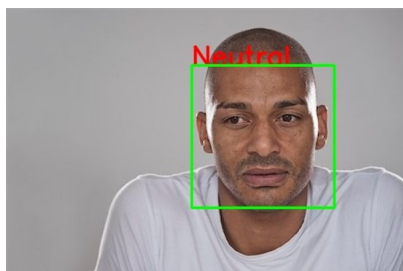


(c) Sad Emotion

Figure 6: Emotion Results 1



(a) Neutral Emotion



(b) Neutral Emotion



(c) Surprise Emotion

Figure 7: Emotion Results 2



The fig. 8 [30] shows different emotions of different people. It can be seen that 4 out of 6 emotions were detected correctly while one image where detected as neutral due minimal changes in the expression of actual emotion while one image was not detected.



Figure 8: Multiple Faces with different emotions

The fig. 9 is a photo from [31] with low lighting. The emotion is detected accurately even though half face in not visible properly, the system catches the parameters such as eyebrows, nose , eyes, mouth shape and accurately detects the emotion.

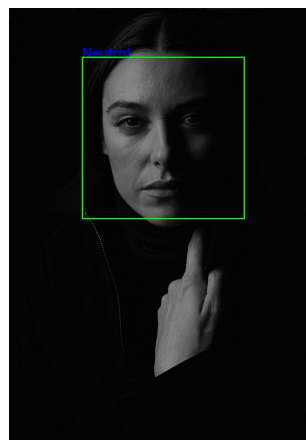


Figure 9: Neutral emotion in Low lighting

The fig. 10 [32] is a photo of a man covering his mouth, even though the mouth is covered the system was able to detect the emotion accurately using the other features such as raised eyebrows, eyes.



Figure 10: Surprise emotion with mouth covering

## 8 Limitations or Challenges

There are numerous challenges faced by this system. The challenges are listed below:

- One of the challenge faced was distinguishing emotion between anger and disgust. Disgust has the least number of image input as a result its accuracy is least, but also the facial expression of anger and disgust is similar like eyebrows lowering.
- Lighting on the face during testing can be a major factor it can cause inaccuracies.
- Detecting emotion for multiple faces at the same time can be done but it can result in delay.
- Covering face while testing can also result in low accuracy. System gives result on the basis of facial expressions like eyebrows, eyes, mouth shape, nose. If the face is covered, the system will only be able to read from the features available during testing.
- Privacy concern is also a major challenge in a system were users face are taken as input.
- Model must be capable of generalizing to novel, unobserved face expressions. It is difficult to achieve this level of generalization because facial expressions varies widely depending on the person, their cultural background, and the situation.

Future of Emotion detection technologies can be enhanced by introducing new methodologies in which face recognition can be done in dim-light and also better bifurcation's between different emotions.

## 9 Conclusions

In conclusion, facial expression recognition is a complex technology that has the potential to revolutionize various fields such as healthcare, marketing, and security. By accurately identifying emotions, it can help doctors diagnose mental health disorders, marketers tailor their advertising campaigns, and security personnel detect potential threats.

However, there are also significant challenges and ethical implications associated with this technology. Developers must overcome technical difficulties such as lighting and pose variations, while also addressing privacy concerns and potential biases in the data used to train these systems.

Despite these challenges, the potential benefits of facial expression recognition are too great to ignore. As we continue to develop and refine this technology, it is important for us to approach it with thoughtfulness and consideration for its impact on society.

## References

- [1] D’Mello, S. K., Kory, J. (2015). "A Review and Meta-Analysis of Multimodal Affect Detection Systems." *ACM Computing Surveys (CSUR)*, 47(3), 1-36.
- [2] Abadi, M., et al. (2016). "TensorFlow: A System for Large-Scale Machine Learning." *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI)*.
- [3] Ekman, P. Friesen, W. (1978). "Facial Action Coding System." Consulting Psychologists Press.
- [4] Yan, W., Garcia, C. (2013). "Facial expression recognition using deep learning." in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.
- [5] 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)*.
- [6] Goodfellow, I., Bengio, Y., Courville, A. (2016). "Deep Learning." MIT Press.
- [7] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "Training algorithm for optimal margin classifiers," *Proc. Fifth Annu. ACM Work. Comput. Learn. Theory*, pp. 144–152, 1992, doi: 10.1145/130385.130401.
- [8] S. J. Wang, H. L. Chen, W. J. Yan, Y. H. Chen, and X. Fu, "Face recognition and micro-expression recognition based on discriminant tensor subspace analysis plus extreme learning machine," *Neural Process. Lett.*, vol. 39, no. 1, pp. 25–43, Feb. 2014, doi: 10.1007/S11063-013-9288-7.

- [9] M. Li, H. Xu, X. Huang, Z. Song, X. Liu, and X. Li, "Facial Expression Recognition with Identity and Emotion Joint Learning," *IEEE Trans. Affect. Comput.*, vol. 12, no. 2, pp. 544–550, Apr. 2021, doi: 10.1109/TAFFC.2018.2880201.
- [10] P. Giannopoulos, I. Perikos, and I. Hatzilygeroudis, "Deep learning approaches for facial emotion recognition: A case study on FER-2013," *Smart Innov. Syst. Technol.*, vol. 85, pp. 1–16, 2018.
- [11] K. Bahreini, R. Nadolski, and W. Westera, "Towards multimodal emotion recognition in e-learning environments," *Interact. Learn. Environ.*, vol. 24, no. 3, pp. 590–605, Apr. 2016,
- [12] C. Clavel, "Surprise and human-agent interactions," *Rev. Cogn. Linguist.*, vol. 13, no. 2, pp. 461–477, Dec. 2015,
- [13] "Adaptive Face Region Representation for Affect Recognition" by Barsoum et al. (2016)
- [14] "Facial Expression Recognition with Intra-Sample Label Switching: A Unified Approach" by E. Barsoum et al. (ICLR 2017)
- [15] IMANO00, "Dataset3Modified." *Kaggle*. [Online]. Available: <https://www.kaggle.com/datasets/imano00/dataset3modified>
- [16] S. Dari, A. A. Hussein, and M. R. El-Sakka, "The Importance of Data Preprocessing in Machine Learning: A Review," in *Procedia Computer Science*, vol. 65, 2015, pp. 1042–1051. doi: 10.1016/j.procs.2015.09.037.
- [17] R. K. Adu-Gyamfi, M. Zhang, S. M. Anthony, "The Benefits of Resizing Datasets for Machine Learning Tasks," in *2019 IEEE International Conference on Big Data (Big Data)*, 2019, pp. 2272–2275. doi: 10.1109/BigData47090.2019.9006418.
- [18] J. Li, Q. Fan, J. Wang, "Benefits of Grayscaleing the Dataset Images for Deep Learning," in *2021 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2021, pp. 1566–1570. doi: 10.1109/ICASSP39728.2021.9413639.
- [19] Yamashita, R., Nishio, M., Do, R.K.G. et al. Convolutional neural networks: an overview and application in radiology. *Insights Imaging* 9, 611–629 (2018).
- [20] Divyanshu Soni., "Translation Invariance in Convolutional Neural Networks", Nov. 2019
- [21] Harsh Yadav., "Dropout in Neural Networks", Jul. 2022
- [22] LeCun, Y., Bengio, Y., Hinton, G. (2015). "Deep learning." *Nature*, 521(7553), 436–444.
- [23] Jason Brownlee., "Gentle Introduction to the Adam Optimization Algorithm for Deep Learning", July. 2017

- [24] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," in Proceedings of the 32nd International Conference on Machine Learning (ICML), 2015, pp. 448-456.
- [25] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning. MIT Press, 2016.
- [26] Christopher M. Bishop. Pattern Recognition and Machine Learning. Springer, 2006.
- [27] Géron, Aurélien. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. O'Reilly Media, 2019. (Chapter 3: Classification)
- [28] Kevin P. Murphy. Machine Learning: A Probabilistic Perspective. MIT Press, 2012. (Chapter 28: Classifier Evaluation)
- [29] Sebastian Raschka and Vahid Mirjalili. Python Machine Learning, 3rd Edition. Packt Publishing, 2019. (Chapter 6: Learning Best Practices for Model Evaluation and Hyperparameter Tuning)
- [30] Lawrence, Kate Campbell, Ruth Skuse, David. (2015). Age, gender, and puberty influence the development of facial emotion recognition. *Frontiers in Psychology*. 6. 10.3389/fpsyg.2015.00761.
- [31] Justin M., Low Light Potrait[Photograph]., Attribution-NonCommercial (CC BY-NC 2.0)
- [32] "image: Freepik.com". This cover has been designed using assets from Freepik.com