Foreword

I feel immense pleasure to foreword the project entitled “Facial Expression Recognition Using Deep Learning” submitted by Utkarsh Sen of School of Statistics, DAVV, Indore.

……………………………

Dr. Vijay Baboo Gupta

**Head of Department**

School of Statistics

DAVV, Indore

CERTIFICATE

The research work embodied in the present project entitled “Facial Expression Recognition Using Deep Learning” has been carried out in the supervision of Dr. Arpita Lakhre by Utkarsh Sen. The work reported herein is original and does not form part of any other project or dissertation based on which a degree or award was conferred on an earlier or to any other scholar.

I understand the University’s policy on plagiarism and declare that the thesis and publications are my own work, expect where specifically acknowledged and has not been copied from other sources or been previously submitted for award or assessment.

Dr. Arpita Lakhre

**Assistant Professor**

School of Statistics,DAVV

Dr. Vijay Baboo Gupta

Head Of Department

School Of Statistics,DAVV

DECLARATION

I, Utkarsh Sen, a student of B.Sc. (Hons) Applied Statistics & Analytics at School of Statistics (Semester-VI), Devi Ahilya Vishwavidyalaya, hereby declare that the project report entitled " Facial Expression Recognition Using Deep Learning " is my original work under the guidance of Dr. Arpita Lakhre.

I affirm that this project has not been submitted elsewhere for any other degree or diploma, and the sources or information used in this report have been duly acknowledged. Any contribution from other individuals or sources has been appropriately credited.

I also declare that the project work complies with the academic standards and guidelines provided by the university.

Utkarsh Sen

**ST4A-2106**

Date: 24/06/2024

Place: Indore

ACKNOWLEDGEMENT

I would like to express my heartfelt gratitude to Dr. Arpita Lakhre, my project guide, for her unwavering support, invaluable guidance, and insightful feedback throughout the course of this project. Her expertise and encouragement have been instrumental in shaping the success of this endeavour.

I am also deeply thankful to Dr. Vijay Baboo Gupta, the head of the department, for providing the necessary resources and a conducive environment for the project to thrive.

My sincere thanks go out to my friends and family for their encouragement, understanding, and motivation during this academic journey. Your unwavering support has been my source of strength.

Last but not least, I extend my appreciation to School of Statistics for fostering an environment of learning and growth, and for providing me with the platform to explore and realize my academic potential. Thank you all for being an integral part of this fulfilling journey.

Utkarsh Sen

**ST4A-2106**

**24/06/2024**

|  |  |  |
| --- | --- | --- |
| **Index** | | |
| **S. no.** | **Topic** | **P.no** |
| 1. | Abstract | **7** |
| 2. | Introduction | **8** |
| 3. | Literature Review | **10** |
| 4. | Background Information | **12** |
| 5. | Methodology | **14** |
| 6. | Future Aspect | **21** |
| 7. | Drawbacks | **23** |
| 8. | Conclusions | **26** |
| 9. | References | **27** |

1. **Abstract**

Automatic facial emotion recognition (FER) has become an increasingly important field in computer vision due to its potential applications in human-computer interaction and social signal processing. Deep learning approaches have achieved remarkable success in FER, surpassing traditional methods. This paper reviews the state-of-the-art deep learning techniques for FER. We discuss the typical pipeline of a FER system, including face detection, preprocessing, feature extraction, and classification. We explore deep learning architectures employed for FER, such as convolutional neural networks (CNNs). The trained model is used in web development. Finally, we discuss the challenges and future directions of deep learning-based FER.

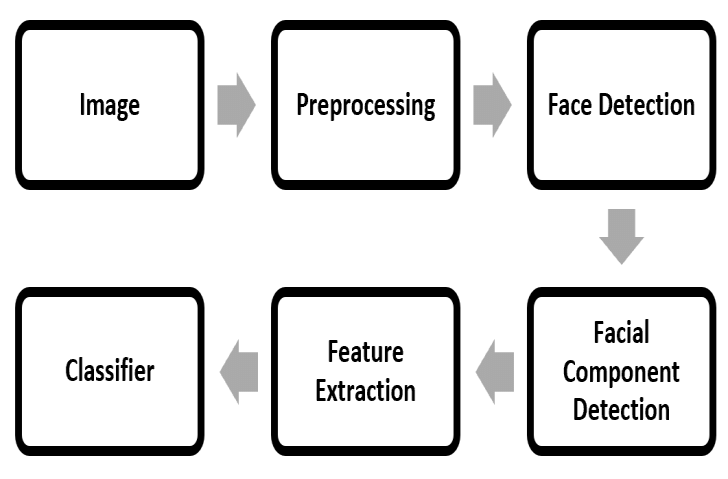
1. **Introduction**

Automatic emotion recognition is a significant research area that addresses two different subjects: psychological human emotion recognition and artificial intelligence (AI). The emotional state of humans can be inferred from verbal and non-verbal information captured by various sensors, for example, from facial changes, tone of voice, and physiological signals. In 1967, Mehrabian showed that 55% of emotional information was visual, 38% vocal, and 7% verbal. Facial changes during communication are the first signs that transmit the emotional state, which is why most researchers are very interested in this modality.

Facial Emotion Recognition (FER) is a flourishing study topic with many breakthroughs in industries such as automatic translation systems and machine-to-human contact. Classical FER consists of two main steps: feature extraction and emotion recognition. Additionally, image preprocessing, including face detection, cropping, and resizing, is crucial. Face detection isolates the facial region after removing the backdrop and non-face areas. Finally, the retrieved characteristics are used to classify emotions, commonly with the help of neural networks (NN) and other machine learning approaches. The challenge of facial emotion recognition is to automatically recognize facial emotion states with high accuracy. This is difficult because the same emotional state can be expressed differently by different people depending on their mood, skin colour, age, and surrounding environment. Generally, FER is separated into the following stages:

1. **Preprocessing**: Detecting an image of a face.
2. **Face Detection**: Detecting facial components from the region.
3. **Feature Extraction**: Extracting informative features from different parts of the face.
4. **Classification**: Training a classifier to generate labels for the emotions using the training data.

Deep learning is a part of machine learning approaches that can be adapted to emotion recognition and facial expression analysis. However, deep learning depends on data size, which may affect its performance.



1. **Literature Review**

Facial emotion recognition (FER) has emerged as a crucial area of computer vision, fuelled by its potential to revolutionize human-computer interaction (HCI) and social signal processing. Traditional methods relied on manually crafted feature extraction, limiting performance and requiring significant domain expertise. However, deep learning approaches have propelled FER forward, achieving remarkable success by enabling automatic feature learning from vast datasets. This review delves into the state-of-the-art deep learning techniques employed for FER.

* 1. **FER System Pipeline**

A typical FER system follows a multi-stage process:

1. **Face Detection**: Identifies and locates human faces within an image or video frame.
2. **Preprocessing**: Images undergo normalization for size consistency, noise reduction, and facial alignment to ensure uniformity.
3. **Feature Extraction**: Deep learning models automatically extract features that capture the emotional information embedded within facial images.
4. **Classification**: The extracted features are classified into distinct emotional categories, ranging from basic emotions (happiness, sadness, anger) to more nuanced emotional states.
   1. **Deep Learning Architectures**

Convolutional Neural Networks (CNNs) are the dominant deep learning architecture for FER tasks. CNNs excel at extracting spatial features from images, making them ideal for capturing the subtle variations in facial expressions that convey emotions. Variants of CNNs, such as VGGNet, ResNet, and Inception, have demonstrated impressive performance in FER applications.

* 1. **Web Development**

While the core of FER lies in research and model development, the ultimate goal is its practical application. Integration of the trained model into web development opens doors for real-world scenarios. Imagine an interactive platform that adapts to a user's emotional state through facial recognition, personalizing the user experience in real-time. By addressing these challenges and actively exploring these promising research avenues, deep learning has the potential to significantly enhance FER accuracy and pave the way for its seamless integration into real-world applications, fundamentally transforming human-computer interaction and social signal processing.

1. **Background Information**
   1. **Emotion Recognition**

Emotion recognition is a branch of computer science that deals with methods and strategies for detecting emotions in facial expressions. It is expected that expressions can be the next communication medium with computers. The majority of this field's research focuses on recognizing human emotions from movies or auditory data. Most of the research has focused on recognizing and matching faces, but convolutional neural networks (CNNs) have been utilized to infuse emotions into photos. Emotion recognition involves identifying emotions and the strategies and procedures used to do so. Emotions can be detected through facial expressions, verbal signals, and other indicators. Machine learning, neural networks, artificial intelligence, and emotional intelligence are just a few methods used to infer emotions. Emotion recognition is gaining traction in research, which is critical to solving various challenges.

* 1. **Facial Emotion Recognition**

Facial emotion recognition is a research area that tries to identify the emotion from human facial expressions. Surveys state that developments in emotion recognition make complex systems simpler. Emotion recognition is a difficult process since emotions can differ depending on the environment, appearance, culture, and facial reaction, resulting in unclear data.

* 1. **Deep Learning**

Deep learning is a machine learning technique that models the data designed to do a particular task. Deep learning in neural networks has wide applications in image recognition, classification, decision making, pattern recognition, etc.

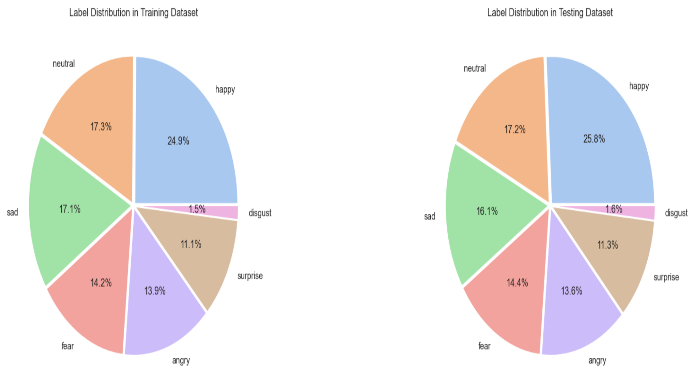
1. **Methodology**

The proposed technique, the emotion database used for the study, and the inception model are explained in this section. This project uses a Haar classifier for human detection. The Haar classifier is trained by Haar-like small features, which are commonly used as texture descriptors. The main features of Haar-like features are linear, edge, centre, and diagonal. The Haar-like feature can reflect the grey-level change of the image, making it effective in explaining the face as many features of external body parts have obvious contrast change characteristics.

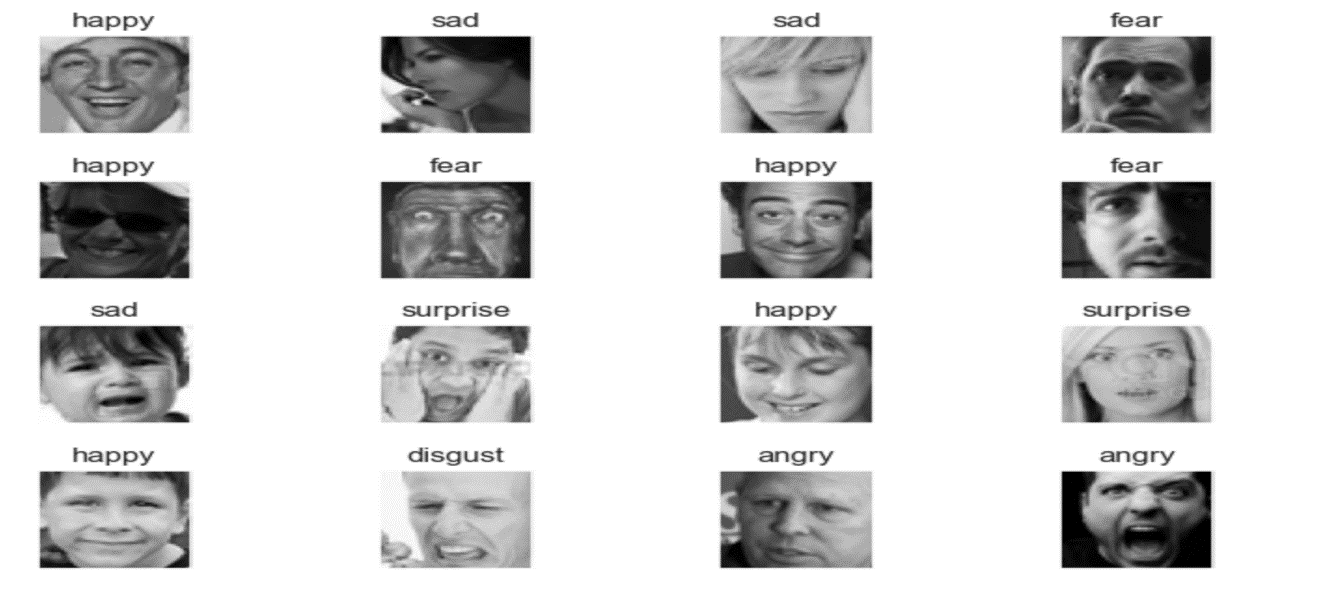
* 1. **Data Description**

The dataset used is the Facial Expression Recognition (FER) dataset. It contains two files: train and test. The train file contains a total of 28,821 images, and the test file has 7,066 images. The images are labeled into the following classes: anger, sadness, happiness, neutral, disgust, fear, surprise.

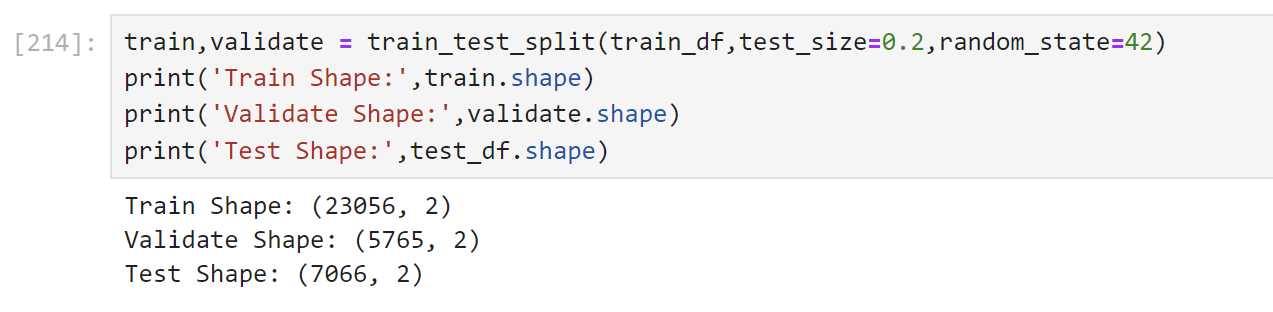
The distribution of various classes is as follows:

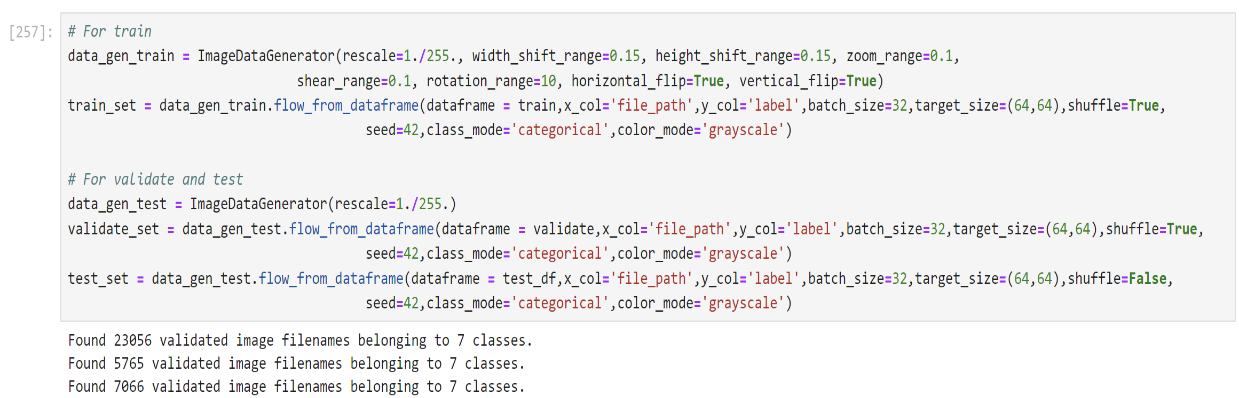


The image in the dataset looks like:



* 1. **Preprocessing**
* **Splitting Data**: The train dataset is further divided into two parts: train containing 23,056 images and validate containing 5,765 images.



* **Data Augmentation**: All three datasets went through data augmentation. The images are standardized, flipped, rotated, and shifted to artificially expand the dataset and improve model generalization.
* 
  1. **Neural Network Design**
     1. **Model Building**

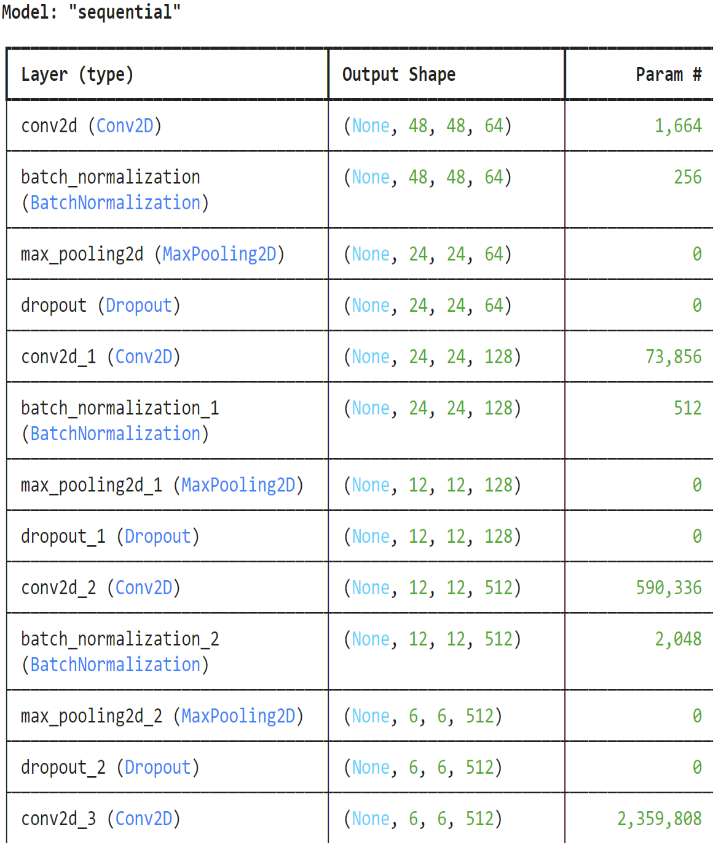
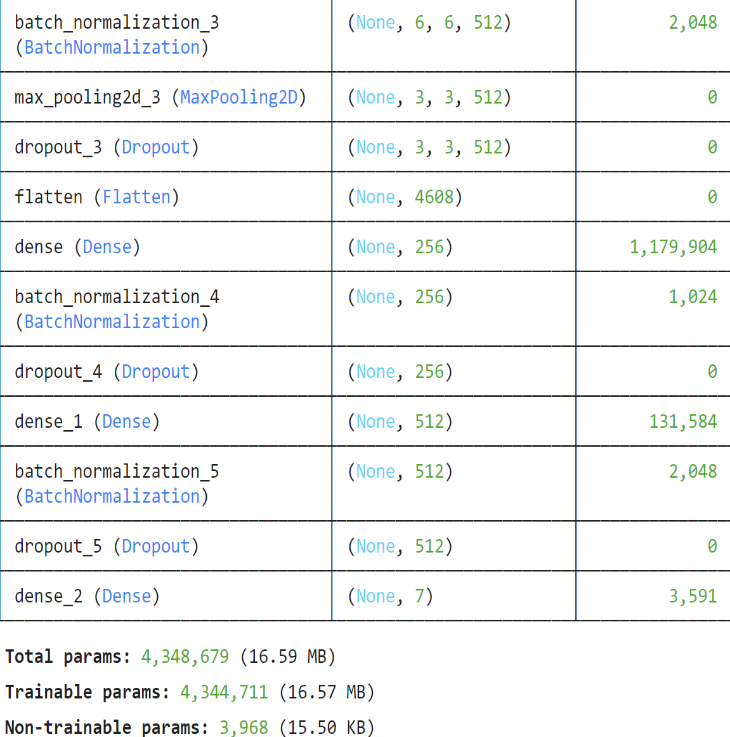
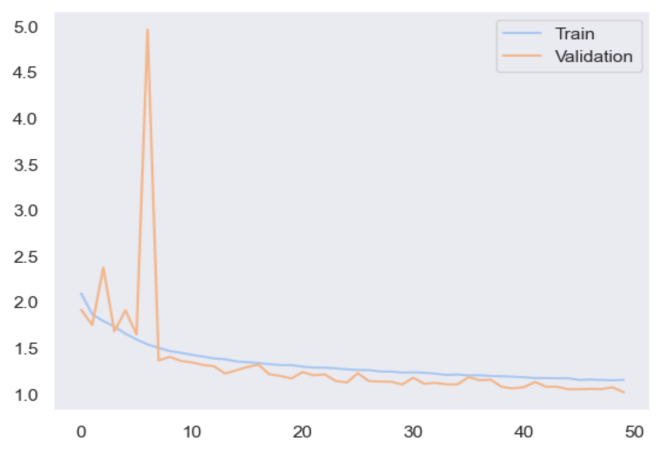
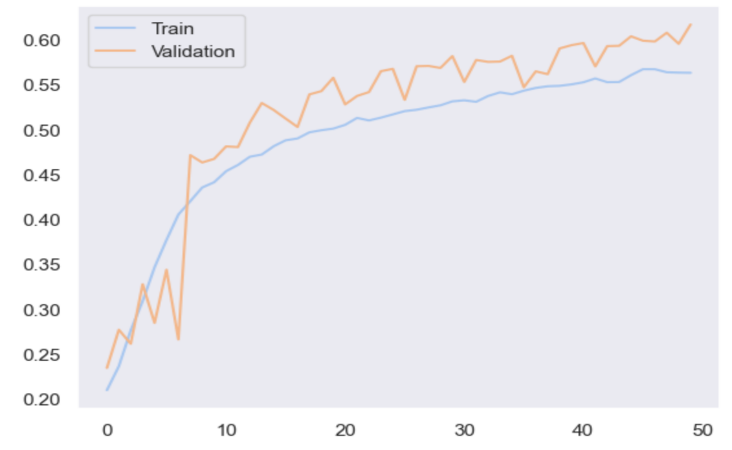
1. **1st Layer**: Convolutional layer with 64 filters of size (5, 5), ReLU activation, and batch normalization. MaxPooling and Dropout layers are added for regularization.
2. **2nd Layer**: Convolutional layer with 128 filters of size (3, 3), ReLU activation, and batch normalization. MaxPooling and Dropout layers are added for regularization.
3. **3rd Layer**: Convolutional layer with 512 filters of size (3, 3), ReLU activation, and batch normalization. MaxPooling and Dropout layers are added for regularization.
4. **4th Layer**: Convolutional layer with 512 filters of size (3, 3), ReLU activation, and batch normalization. MaxPooling and Dropout layers are added for regularization.
5. **Flatten Layer**: Flatten the output from the convolutional layers to be fed into the fully connected layers.
6. **Fully Connected Layer 1**: Dense layer with 256 units and ReLU activation, followed by batch normalization and dropout for regularization.
7. **Fully Connected Layer 2**: Dense layer with 512 units and ReLU activation, followed by batch normalization and dropout for regularization.
8. **Output Layer**: Dense layer with softmax activation for multi-class classification, with the number of units equal to the number of classes in the dataset.
9. **Compilation**: The model is compiled with the Adam optimizer, categorical cross-entropy loss function, and accuracy metric.

fig.(1) fig.(2)

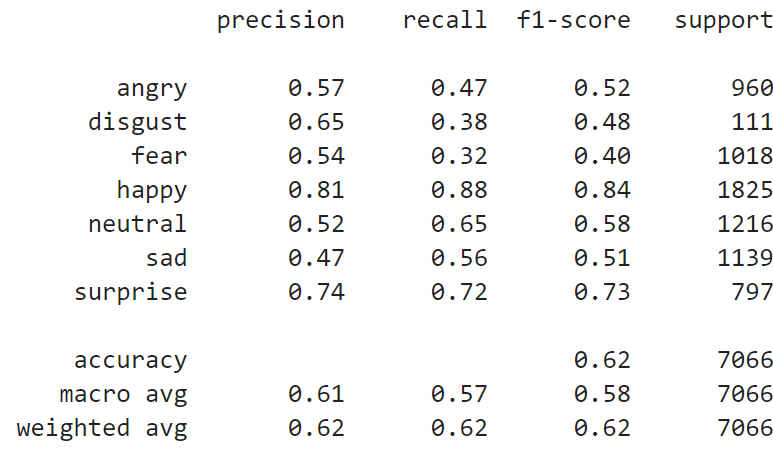
* + 1. **Training**

The model was trained for 50 epochs with a train set of 23,056 images and a validate set of 5,765 images.

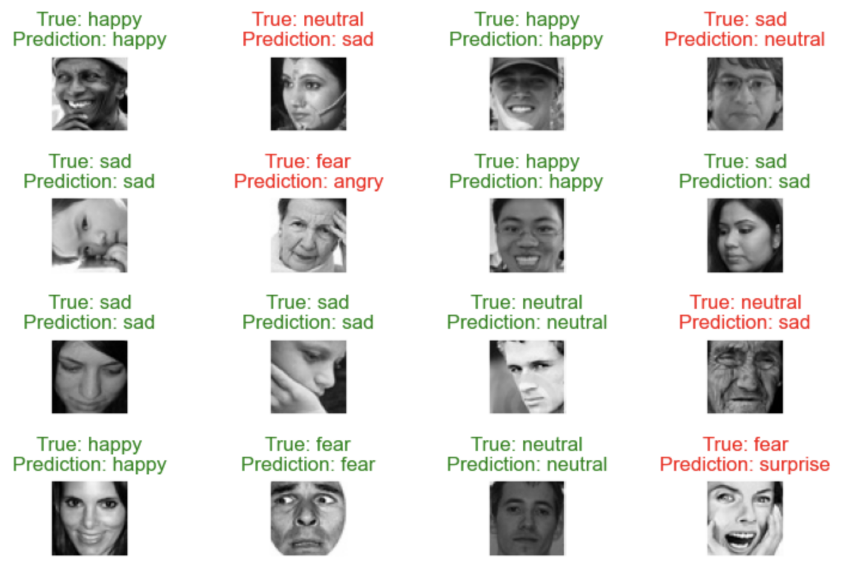
* + 1. **Model Evaluation**
* **Loss**: After 50 epochs, train loss was 1.15 and validate loss was 1.01. The graph shows both train loss and validate loss decrease with time, indicating the model is not overfitting.
* **Accuracy**: Both train and validate accuracy increase with epochs. The train set accuracy at the last epoch is 56.28%, and the validate set accuracy is 61.67%.

Loss graph Accuracy graph

* + 1. **Testing**

The test set has a loss of 1.01 and accuracy of 62.5%.

Classification report:



Prediction on test data looks like:

* 1. **Model Deployment**
     1. **Face Detection**

Face detection is a pre-processing phase to recognize the facial expressions of humans. An image is segmented into two parts: faces and other non-face regions. Numerous methods are used for face detection. The first step in face detection involves loading the image and converting it into grayscale. A Haar cascade classifier then runs through this gray image to detect the face. The detected face is standardized, and the standardized image is used for prediction.

Code:

Conversion of gray image Face detection

Output:





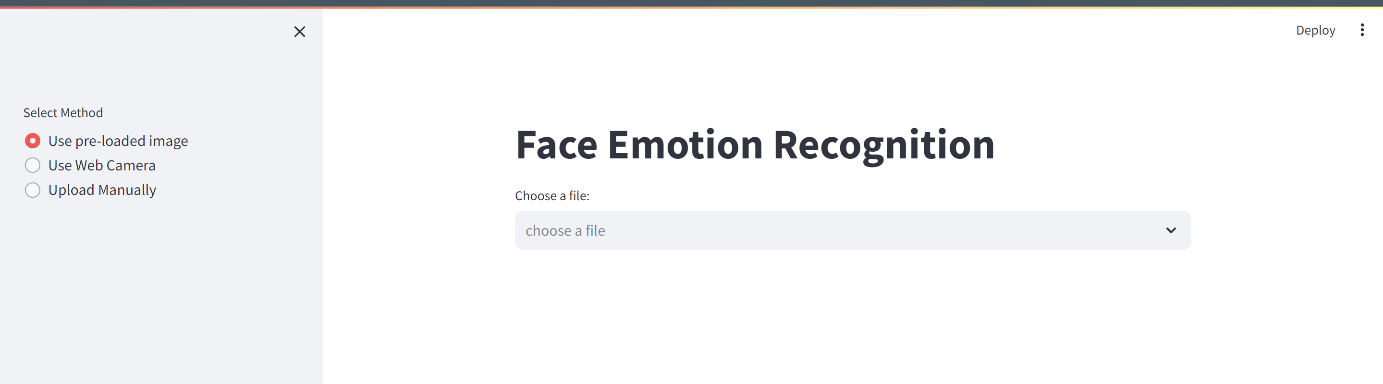
Gray image Detected face

* + 1. **Model Deployment**

After detecting the face, the saved model is used to make predictions. The front-end provides three options to load an image:

1. **Pre-loaded Images**: Use pre-loaded images saved on the cloud.
2. **Web Camera**: Use a web camera to capture a live image.
3. **Upload Image**: Upload an image from your device.

The result is displayed as a single word indicating the detected emotion. A detailed view shows the probability percentage of different emotions.



1. **Future Aspects**

Despite the advancements in FER, several challenges and opportunities for future research remain:

1. **Data Diversity and Augmentation**: Expanding the dataset to include a more diverse range of faces, emotions, and cultural backgrounds can improve the model's robustness and generalizability. Advanced data augmentation techniques can further enhance the training process.
2. **Real-Time Processing**: Improving the efficiency of real-time emotion recognition systems to reduce latency and enhance user experience is crucial. Optimizing deep learning models for faster inference without compromising accuracy is a key area of focus.
3. **Multi-Modal Emotion Recognition**: Integrating other modalities, such as voice and physiological signals, can provide a more comprehensive understanding of human emotions. Combining facial expressions with vocal tone and other signals can lead to more accurate and nuanced emotion recognition.
4. **Addressing Bias**: Ensuring that FER systems are free from biases related to age, gender, ethnicity, and other factors is essential for fair and ethical applications. Developing techniques to detect and mitigate biases in training data and models is a critical area of research.
5. **Robustness to Environmental Factors**: Enhancing the robustness of FER systems to varying lighting conditions, occlusions, and different environments can improve their practical applicability in real-world scenarios.
6. **Emotion Detection in Dynamic Scenes**: Extending FER to video sequences to recognize emotions in dynamic scenes and interactions is a promising direction. Temporal models, such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks, can be explored for this purpose.
7. **User Privacy and Ethical Considerations**: As FER systems become more prevalent, addressing user privacy and ethical concerns is paramount. Implementing secure data handling practices and ensuring user consent and transparency are vital for the responsible deployment of FER technologies.

By addressing these challenges and exploring these research avenues, deep learning-based FER can achieve higher accuracy and broader applicability, fundamentally transforming human-computer interaction and social signal processing.

## ****Drawbacks****

While the deep learning-based Facial Emotion Recognition (FER) system described in this report demonstrates significant advancements and potential, there are several drawbacks and limitations that need to be addressed for further improvement and broader applicability.

### ****1. Data Limitations****

**Imbalanced Datasets**: The dataset used for training the model may have an imbalanced distribution of emotions, leading to biased performance where the model is more accurate for certain emotions and less accurate for others.

**Lack of Diversity**: The datasets often lack diversity in terms of age, ethnicity, and cultural backgrounds, which can result in models that do not generalize well across different populations.

### ****2. Real-Time Constraints****

**Latency Issues**: Real-time emotion recognition systems require high processing speeds, and the current model may experience latency issues, especially when integrated into web applications, affecting user experience.

**Resource Intensive**: Deep learning models, particularly CNNs, are computationally intensive, requiring significant hardware resources for both training and inference, which can be a limitation for deployment on devices with limited processing power.

### ****3. Environmental Sensitivity****

**Lighting and Occlusions**: The performance of the FER system can be adversely affected by varying lighting conditions, shadows, and occlusions (e.g., glasses, hats, masks), which are common in real-world scenarios.

**Background Noise**: In practical applications, background noise and other non-facial elements in images can interfere with face detection and emotion recognition accuracy.

### ****4. Model Interpretability****

**Black Box Nature**: Deep learning models, including CNNs, are often considered black boxes due to their complex architectures, making it difficult to interpret how they make specific predictions. This lack of interpretability can be a drawback in understanding and trusting the model’s decisions.

### ****5. Generalization Issues****

**Overfitting**: Despite using techniques like dropout and data augmentation, deep learning models can still overfit to the training data, resulting in decreased performance on unseen data.

**Domain Adaptation**: The model trained on a specific dataset might not perform well when applied to a different dataset or real-world images, indicating poor domain adaptation.

### ****6. Privacy and Ethical Concerns****

**Privacy Invasion**: The deployment of FER systems raises significant privacy concerns as it involves the continuous monitoring and analysis of individuals’ facial expressions, potentially without their consent.

**Ethical Implications**: There are ethical considerations regarding the use of FER systems, especially in sensitive applications such as surveillance, where misuse could lead to discrimination or biased decision-making.

### ****7. Limited Emotion Categories****

**Complex Emotions**: The current model focuses on basic emotions such as anger, sadness, happiness, etc. Recognizing complex or mixed emotions, which are common in human interactions, remains a challenge.

**Context Ignorance**: FER systems often ignore the context in which an emotion is expressed, which is crucial for accurately interpreting emotions. For instance, a smile might indicate happiness in one context and sarcasm in another.

### ****8. Dependency on Large Datasets****

**Data-Hungry Nature**: Deep learning models require large amounts of labeled data for training. Acquiring and annotating such datasets is time-consuming and costly, which can be a barrier to developing robust FER systems.

### ****9. Bias and Fairness****

**Algorithmic Bias**: The model can inherit biases present in the training data, leading to unfair treatment of certain demographic groups. Ensuring fairness and mitigating bias is an ongoing challenge.

1. **Conclusion**

Facial Emotion Recognition (FER) using deep learning has shown significant promise and success in various applications, from human-computer interaction to social signal processing. This study reviewed the FER pipeline, including face detection, preprocessing, feature extraction, and classification, with a focus on Convolutional Neural Networks (CNNs) as the primary deep learning architecture. The model developed in this study demonstrated reasonable accuracy and performance, highlighting the effectiveness of CNNs in capturing and classifying facial emotions.

The practical application of the FER model was illustrated through its integration into a web development platform, offering real-time emotion recognition from pre-loaded images, live webcam feeds, or user-uploaded images. This integration paves the way for interactive and personalized user experiences, adapting to users' emotional states.

1. **References**

 **Dataset**

* Facial Expression Recognition Dataset (2013). Available at: https://www.kaggle.com/datasets/jonathanoheix/face-expression-recognition-dataset

 **Research Paper**

* Pandey, A., & Gupta, A. (2022). Facial Emotion Detection and Recognition. Available at: https://www.researchgate.net/publication/361108119\_FACIAL\_EMOTION\_DETECTION\_AND\_RECOGNITION

 **Website**

* Dive Deep into Deep Learning. Available at: [www.d2l.ai](http://www.d2l.ai)

 **Book**

* Bhatia, M. (2012). Facial Expression Recognition System. Published in July 2012.

 **Streamlit**

* Streamlit. Available at: [www.streamlit.io](http://www.streamlit.io)

 **Book**

* Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press. Available at: [www.deeplearningbook.org](http://www.deeplearningbook.org)

 **Journal Article**

* Li, S., & Deng, W. (2020). Deep Facial Expression Recognition: A Survey. IEEE Transactions on Affective Computing, 12(2), 119-135. Available at: [IEEE Xplore](https://ieeexplore.ieee.org/document/8956785)

 **Conference Paper**

* Zhang, Z., Luo, P., Loy, C. C., & Tang, X. (2014). Facial Landmark Detection by Deep Multi-task Learning. In European Conference on Computer Vision (ECCV). Available at: [SpringerLink](https://link.springer.com/chapter/10.1007/978-3-319-10599-4_7)

 **Blog**

* Brownlee, J. (2019). A Gentle Introduction to Convolutional Neural Networks. Available at: Machine Learning Mastery

 **Online Course**

* Ng, A. (2020). Convolutional Neural Networks. Available at: [Coursera](https://www.coursera.org/learn/convolutional-neural-networks)