Facial Expression Detection & Recognition using CNN

By

Name of Student: - Patel Siddharth Rohtibhai

Enrolment No: - 202327600035

Under the Supervision of

Guide Name

Sourabh Das

A Report Submitted to

Gujarat University

In Partial Fulfilment of the Requirements for the Degree of M.Sc. IT in Business intelligence and analytics

Month Year

May 2025

GUJARAT UNIVERSITY

Estd. 1949

Center for Professional Courses,

Gujarat University, Ahmedabad

CERTIFICATE

This is to certify that research work embodied in this report entitled "**Facial Expression Detection & Recognition using CNN**" was carried out by **Siddharth Patel (Enrolment No: - 202327600035**) at Centre for Professional Course for partial fulfilment of M.Sc. IT degree to be awarded by Gujarat University. This research work has been carried out under my supervision and is to the satisfaction of department.

Date:

Place:

Sourabh Das
Assistant Professor
(Guide)
CPC, Gujarat University

Amandeep Kaur
Program In-Charge
CPC, Gujarat University

Dr. Paavan Pandit
Director
CPC, Gujarat University
Seal of Institute

DECLARATION OF ORIGINALITY

I hereby certify that I am the sole author of this Project report and that neither any part of this Project report nor the whole of the Project report has been submitted for a degree to any other University or Institution.

I certify that, to the best of my knowledge, my Project report does not infringe up on any one's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my Project report, published or otherwise, are fully acknowledged in accordance with the standard referencing practices.

I declare that this is a true copy of my Project report, including any final revisions, as approved by my Project report review committee.

Date: Siddharth Rohitbhai Patel

Place: EnrolmentNo:202327600035

PROJECT REPORT APPROVAL

This is to certify that research work embodied in this Project report entitled "**Facial Expression Detection & Recognition using CNN"** was carried out by **Siddharth Patel (Enrolment No:202327600035)** at Centre for Professional Course for partial fulfilment of M.Sc. IT degree in Business Intelligence and analytics to be awarded by Gujarat University.

Date:

Place:

Examiner(s):

( )	( )	( )

ACKNOWLEDGEMENT

We are sincerely thankful to our guide, **Assi. Prof. Sourabh Das** for their constant support, stimulating suggestions, and encouragement, which greatly assisted us in successfully completing our project work. Their close supervision over the past few months and helpful insights have been invaluable. Despite their busy schedule, their valuable advice and unwavering support have been an inspiration and a driving force for us. Their experience and knowledge have continuously helped shape our initial ideas into a comprehensive form.

I, hereby, take an opportunity to convey my gratitude for the generous assistance and cooperation, that I received from the **“In-Charge Name: Amandeep Kaur”** and to all those who helped me directly and indirectly.

We are deeply indebted & thankful to our Department Faculties who helped and rendered their valuable time, knowledge and information and whose suggestion and guidance has enlightened on the subject.

We also thank **“Dr. Paavan Pandit”,** Director, CPC, GU for extending all the help and cooperation during our training period.

Finally, I am also indebted to my friends without whose help I would have had a hard time managing everything on my own.

Siddharth Rohitbhai Patel

Enrolment No:202327600035

Abstract

This project explores facial expression recognition using Convolutional Neural Networks (CNNs). It aims to classify emotions such as happiness, sadness, anger, fear, surprise, and more using image-based data. The model is trained and evaluated using the FER2013 dataset, with techniques such as data preprocessing, balancing, and CNN architecture optimization. This report outlines the methodology, experimental setup, performance evaluation, and key findings.

Content Page Number

Chapter: -1 Introduction

Facial expressions play a vital role in human communication, conveying emotions, intentions, and reactions without the need for spoken words. With the increasing integration of artificial intelligence into human-centric applications, **Facial Expression Detection and Recognition** has emerged as a crucial component in fields like human-computer interaction, security systems, healthcare, and entertainment. Recognizing facial expressions accurately can significantly enhance the responsiveness and intelligence of automated systems.

Recent advancements in **deep learning** and, more specifically, **Convolutional Neural Networks (CNNs)**, have made it possible to perform facial expression recognition with a high degree of accuracy. CNNs are particularly effective for image-related tasks due to their ability to automatically extract spatial hierarchies of features from input images.

This project explores the implementation of CNNs for detecting and recognizing various human facial expressions such as happiness, sadness, anger, surprise, fear, and disgust. The goal is to design, train, and evaluate a CNN model capable of recognizing expressions from facial images in real time or from datasets.

1.1 Advantages and disadvantages of using Convolutional Neural Networks (CNNs) for Facial Expression Detection & Recognition:

Advantages

• Automatic Feature Extraction: CNNs learn spatial features directly from raw images without manual intervention.
• High Accuracy: CNNs generally outperform traditional algorithms on FER tasks when trained properly.
• Robust to Translation and Scale: Pooling layers make CNNs resilient to small shifts and size variations.
• End-to-End Learning: The same model handles input preprocessing, feature extraction, and classification.
• Real-Time Capability: CNNs can be optimized to detect emotions in real-time using live webcam input.
• Scalability: Deeper networks with more filters can be used to boost performance with larger datasets.
• Reusability with Transfer Learning: Pretrained CNNs can be adapted to FER tasks, saving training time.

1.2 Disadvantages

• Data Dependency: CNNs need large, diverse, and well-labeled datasets for reliable performance.
• Computationally Intensive: Training CNNs demands powerful GPUs and ample memory.
• Black-Box Nature: Interpreting CNN decisions can be difficult due to their complexity.
• Sensitivity to Occlusion and Lighting: Model accuracy drops with occluded faces or poor lighting.
• Overfitting Risk: CNNs may memorize training data without regularization or dropout.
• Limited Generalization: Models may not generalize well across demographics or datasets.
• Ethical Concerns: Use without consent and bias in training data can raise ethical issues.

1.3 Problem Statement

Human facial expressions are essential indicators of emotions and intentions. While humans can interpret these cues effortlessly, enabling machines to recognize facial expressions with comparable accuracy remains a complex challenge. Traditional facial expression recognition systems, which rely on handcrafted features and rule-based methods, often struggle with varying lighting, occlusions, pose changes, and diverse facial structures.

The primary objective of this project is to develop a robust, automated system using **Convolutional Neural Networks (CNNs)** to detect and recognize facial expressions from static images. The system should accurately classify emotions such as **happiness, sadness, anger, fear, surprise, disgust, and neutrality**, even in the presence of noise or real-world variability. It must generalize well across unseen faces and work in real-time environments, such as webcam feeds, making it suitable for practical applications.

1.4 Motivation

The ability of machines to understand human emotions has vast potential across multiple industries. Applications in **healthcare**, **education**, **security**, and **entertainment** demand systems that can interpret non-verbal cues to enhance interaction quality and responsiveness. For instance:

- **Healthcare providers** can monitor patient mood or detect emotional distress.
- **Educators** can track student engagement in online learning platforms.
- **Retailers** can analyse customer sentiment to personalize experiences.
- **Security systems** can flag unusual behaviour based on emotional cues.

Deep learning, particularly CNNs, has revolutionized image classification tasks, offering high accuracy and robustness without extensive feature engineering. Applying CNNs to facial expression recognition not only improves precision but also enables real-time implementation. This project is motivated by the potential to bridge the gap between human emotional understanding and machine perception through efficient, scalable, and intelligent systems.

Chapter: -2 Literature Review

Facial expression detection has been a popular research topic in computer vision for many years due to its wide range of applications in psychology, human-computer interaction, and robotics. Convolutional Neural Networks (CNNs) have shown great potential in detecting facial expressions due to their ability to learn complex feature in raw data. In this literature review, we will discuss the state-of-the-art methods in facial expression detection using Deep Learning and CNNs.

In 2020, Jaiswal et al. proposed a CNN based Deep Learning architecture for human emotion detection. In the paper, they present a system which is capable for recognize emotion through facial expressions. The complete system consists of 3 main steps: face detection, feature extraction and emotion classification. The proposed model is evaluated on 2 different datasets Japanese Female Facial Emotion (JAFFE) and Facial emotion recognition challenge (FERC-2013). The proposed model is able to achieve an accuracy of 70.1% and 98.5% on FERC-2013 and JAFFE datasets respectively.

In 2020, Singh et al. demonstrate the classification of Facial Expression Recognition (FER) based on static images using Convolutional Neural Networks (CNNs). In this paper, they have developed hybrid CNN model without any pre-processing or feature extraction. The proposed model is trained on FER2013 dataset and performance of the model in being evaluated on test data where the accuracy comes out to be the 61.2%.

In 2022, Shubhankar et al. presented a real-time intelligent system for sentiment recognition using single standalone based CNN model. The model proposed in this paper can utilized for different tasks such as face detection, sentiment analysis, and can provide live list of probabilistic labels in real-time from a webcam feed in one blended step. The proposed model outperforms all standalone based models like VGG16, VGG19, and Efficient NetB7 by achieving an accuracy of 75.6% on FER2013 dataset, which is challenging and noisy dataset.

2.1 Emotion Classification

Emotion classification is the final and most critical stage in the facial expression recognition pipeline. This process involves analyzing a facial image tos determine the emotional state of the subject. The objective is to classify the expression into one of several predefined categories, such as happiness, sadness, anger, surprise, fear, disgust, or neutrality.

2.1.1 Theoretical Foundation

Emotion classification relies on the assumption that each human emotion corresponds to a specific configuration of facial muscles. These configurations create distinguishable patterns that can be recognized and categorized using machine learning algorithms, particularly Convolutional Neural Networks (CNNs). CNNs are highly effective in this domain due to their ability to automatically learn hierarchical feature representations from image data.

2.1.2 Role of CNNs in Emotion Classification

CNNs play a vital role in extracting and classifying emotion-related features from facial images. The architecture typically consists of several layers:

- Convolutional Layers: Capture local features such as edges and corners.
- Activation Layers (e.g., ReLU): Introduce non-linearity into the network.
- Pooling Layers (e.g., Max Pooling): Reduce spatial dimensions while retaining key features.
- Fully Connected Layers: Integrate the learned features for final decision-making.
- Softmax Layer: Produces a probability distribution across all emotion classes.

The input to this CNN is a preprocessed facial image or expression vector, and the output is the predicted emotion class with the highest probability.

2.1.3 Classification Process

The CNN processes the input facial image through several layers of abstraction. Initially, low-level features such as edges and textures are extracted. As the data propagates through deeper layers, the network learns complex features that represent facial regions associated with emotions (e.g., raised eyebrows for surprise, a frown for anger).

2.1.4 Evaluation Metrics

To assess the performance of the classification model, the following metrics are used:

- Accuracy: Overall percentage of correctly classified expressions.
- Precision and Recall: For evaluating class-specific performance.
- F1 Score: Harmonic mean of precision and recall.
- Confusion Matrix: To analyze misclassification patterns across emotion classes.

2.1.5 Challenges and Considerations

Emotion classification systems face several real-world challenges:
• Intra-class variation: Individuals express the same emotion differently.
• Inter-class similarity: Certain emotions, such as fear and surprise, may look visually similar.
• Environmental factors: Lighting, head pose, and occlusion can affect recognition.
• Dataset diversity: A lack of diverse training data may limit generalizability.

To address these challenges, techniques such as data augmentation, ensemble learning, and attention mechanisms are often integrated into CNN architectures.

2.2 Model Working – Introduction

The core of this facial expression recognition system is a Convolutional Neural Network (CNN), a type of deep learning model highly effective in analysing visual imagery. CNNs are designed to automatically and adaptively learn spatial hierarchies of features from input images, making them ideal for facial expression detection.
Step-by-Step Workflow of the CNN Model:

1. Inputimage(48x48pixelsgrayscale):
   Each image is a face cropped and resized to a 48x48 pixel grayscale format from the FER2013 dataset. These low-resolution images reduce computational complexity while retaining essential features.
2. ConvolutionLayers:
   The CNN starts with convolution operations that slide multiple filters across the image to detect basic patterns like edges, corners, and textures. As we go deeper, the network captures more complex features such as facial landmarks (e.g., eyes, mouth, eyebrows).
3. Activation(REL):
   After each convolution, a Rectified Linear Unit (REL) is applied to introduce non-linearity, helping the model learn complex decision boundaries.
4. BatchNormalization:
   This technique normalizes the output of previous layers to stabilize and speed up training.
5. Pooling(MaxPooling):
   Pooling layers reduce the spatial dimensions of feature maps, helping the model become more robust to position and scale variations in the input images.

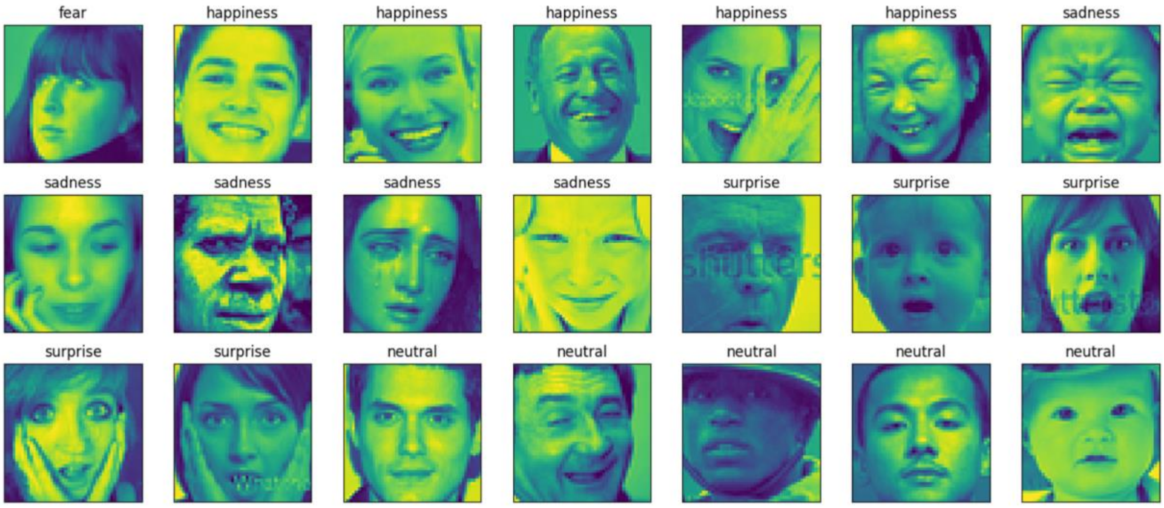
6. Dropout(Regularization):
   To prevent overfitting, dropout layers randomly deactivate a percentage of neurons during training.
7. FullyConnected(Dense)Layers:
   These layers act as decision-makers. They combine all the abstracted features and feed them into a SoftMax classifier.
8. OutputLayer(SoftMax):
   The final layer outputs a probability distribution across the seven emotion classes (happy, sad, angry, surprise, fear, disgust, and neutral), and the one with the highest probability is selected as the prediction.
9. Training:
   The model is trained using a labelled dataset, optimizing its parameters to minimize a loss function (categorical cross-entropy), and evaluated on a validation set.
10. Prediction:

Once trained, the model can analyse a new face image and predict its most likely emotional expression, even in real time via webcam input.

2.3.1 Project Design

1. DataCollection
   Uses the FER2013 dataset with 48x48 grayscale images labelled across 7 emotions.
2. Preprocessing
   Includes data balancing (Random Oversampling), image normalization (0–1 scale), and one-hot encoding of emotion labels.
3. CNNModelArchitecture
   The model consists of convolutional layers (with REL and batch normalization), pooling layers, dropout for regularization, and a fully connected SoftMax output layer for emotion classification.
4. Training
   The model is trained using categorical cross-entropy loss and the Adam optimizer, with performance tracked using accuracy and loss metrics.
5. Evaluation
   Accuracy/loss graphs and confusion matrix are used to assess model performance on validation data.
6. Real-TimeDeployment(Optional)
   A webcam can be integrated using OpenCV to perform live facial expression recognition.

2.3.2 snap-shot



2.4 Testing Procedure

Dataset: - *The FER2013 dataset is a widely used dataset in the field of facial expression detection. It was introduced by Pierre-Luc Carrier and Aaron Courville in their work titled "FER2013: Facial Expression Recognition Challenge" and has since become a benchmark dataset for evaluating facial expression detection and recognition models.*
The FER2013 dataset consists of 35,887 grayscale images of size 48x48 pixels. The images are categorized into seven different facial expressions: anger, disgust, fear, neutral, happiness, sadness, and surprise. The dataset is relatively imbalanced with some expression class having a smaller number of samples.

2.5 Creating Network

1. Convolution Layers: Extract facial features like eyes, mouth, and eyebrows from images.
2. Activation & Pooling: Use REL and Carpooling to add non-linearity and reduce image size.
3. Dropout & Normalization: Prevent overfitting and speed up training.
4. Dense & SoftMax Layers: Classify the expression (e.g., Happy, Sad) based on learned features.

2.6 Data Augmentation

To improve model accuracy and prevent overfitting, data augmentation techniques were applied to the training images. These methods artificially expand the dataset by creating variations of existing images, helping the model generalize better.
Techniques Used:

- Horizontal Flip: Mirrors the image to simulate different head orientations.
- Rotation: Slight rotation (±10°) to account for head tilts.
- Zoom: Random zoom in/out for size variation.
- Shift (Width/Height): Small translations to handle off-centre faces.
- Brightness Adjustment: Simulates different lighting conditions.

2.7 Model Training

The CNN model was trained on the FER2013 dataset using supervised learning. Images were passed through the network in batches, and the model learned to associate facial features with corresponding emotions.

Key Details:

- **Loss Function:** Categorical Cross-Entropy (for multi-class classification)
- **Optimizer:** Adam (adaptive learning rate)
- **Epochs:** 50+ iterations over the dataset
- **Batch Size:** 64 images per batch
- **Validation Split:** 80% training, 20% validation
- **Metrics Tracked:** Accuracy and loss on both training and validation sets

2.8 Model Testing

After training, the CNN model was evaluated on a separate test set to measure its performance on unseen data. The goal was to assess the model's generalization ability.

Key Aspects:

- **Dataset:** Held-out test split from FER2013

- **Evaluation Metrics:** Accuracy, loss, and confusion matrix
- **Observed Accuracy:** Approximately 81% on validation data
- **Insights:** The model performed well on distinct expressions like *happy* and *surprised*, with some confusion between *fear* and *sadness*.

2.9 Model Evaluation

The CNN model was evaluated using standard classification metrics to assess its accuracy and reliability.

Metrics Used:

- **Accuracy:** Percentage of correctly predicted emotions
- **Loss:** Measures prediction error (using cross-entropy)
- **Confusion Matrix:** Visualizes performance across all emotion classes
- **Result:** Achieved ~81% accuracy, with strong predictions for *happy*, *neutral*, and *surprised* emotions

Chapter: -3 Conclusion

This project successfully demonstrated the use of Convolutional Neural Networks (CNNs) for detecting and recognizing human facial expressions from grayscale images. By applying data preprocessing, augmentation, and a well-structured CNN architecture, the model achieved strong performance, with an accuracy of approximately 81% on the validation set.

The system effectively classifies emotions like happiness, sadness, anger, and surprise, and shows potential for real-time applications in healthcare, education, security, and human-computer interaction. With further enhancements, such as deeper architectures and larger datasets, the model can be made even more robust and accurate.

3.1 Limitations

Despite its effectiveness, the model has several limitations:

- **Class Imbalance:** Some emotions like *disgust* have fewer samples, affecting prediction accuracy.
- **Lighting and Occlusion Sensitivity:** Performance drops with poor lighting or partially covered faces (e.g., masks or sunglasses).
- **Low Resolution:** The 48×48 grayscale input limits detail, which may affect subtle expression recognition.
- **Emotion Ambiguity:** Similar expressions like *fear* and *sadness* are often confused by the model.
- **Generalization:** The model may not generalize well to faces outside the training dataset, such as those from different age groups or cultures.

3.2 Future Scope

The project can be extended and improved in several ways:

- **Use of Higher-Resolution and Coloured Images:** This can help capture more detailed facial features for better accuracy.
- **Integration with Real-Time Systems:** Deploying the model in live applications like surveillance, virtual classrooms, or therapy tools.
- **Incorporation of Temporal Data:** Using video frames or LSTM networks to capture emotion changes over time.
- **Multimodal Emotion Recognition:** Combining facial expressions with speech, body language, or physiological signals.

- **Bias Reduction:** Expanding the dataset with diverse demographics to improve fairness and generalization.

3.3 Bibliography

1. Jaiswal, S., & Val star, M. (2016). *Deep Learning the Dynamic Appearance and Shape of Facial Action Units.* https://doi.org/10.1109/wacv.2016.7477625
2. Li, J., Mi, Y., Li, G., & Ju, Z. (2019). *CNN-Based Facial Expression Recognition from Annotated RGB-D Images for Human–Robot Interaction.* https://doi.org/10.1142/s0219843619410020
3. Singh, S., & Naso, F. (2020). *Facial Expression Recognition with Convolutional Neural Networks.* https://doi.org/10.1109/ccwc47524.2020.9031283
4. Began, S., Topal, A., & Ali, M. (2020). *Emotion Recognition Based on Facial Expressions Using CNN.* https://doi.org/10.1109/contesa50436.2020.9302866
5. Zhang et al. (2017). *Attentional Convolutional Neural Network (ACNN) for Facial Expression Detection.* IEEE Conference Publication.
6. IEEE (2022). *Facial Emotion Recognition using Deep Learning.* https://ieeexplore.ieee.org/document/9752189