FACE EMOTION RECOGNITION USING DEEP LEARNING

Minor Project Report (Condensed – Presentation Version)

Submitted by: Mohd Anas
Course: M.Sc. Artificial Intelligence & Machine Learning
Semester: III (2025–26)
Supervisor: Prof. Jahiruddin
Department of Computer Science
Jamia Millia Islamia, New Delhi

Abstract

Emotion recognition plays a vital role in modern Artificial Intelligence systems by enabling machines to understand human emotional states through facial expressions. Facial Emotion Recognition (FER) has wide applications in education, healthcare, security, human–computer interaction, and psychological analysis.

This project presents a **deep learning–based Face Emotion Recognition system** using **Convolutional Neural Networks (CNNs)** and **ConvNeXt transfer learning architecture**, trained on the **FER2013+ dataset**. The dataset consists of **48×48 grayscale facial images** categorized into **eight emotions**: Angry, Contempt, Disgust, Fear, Happy, Neutral, Sad, and Surprise.

The system integrates **MTCNN** for accurate **multi-face detection** and is deployed as a **real-time Streamlit web application** that supports image upload and webcam-based emotion recognition. Experimental results show that the fine-tuned ConvNeXt model achieves **~79% validation accuracy**, with strong performance for emotions such as Happy and Neutral.

This project demonstrates the effectiveness of deep learning in automatically extracting facial features and highlights the feasibility of deploying real-time emotion-aware intelligent systems.

Introduction

Human emotions are fundamental to communication and social interaction. Facial expressions provide a natural way to convey emotional states such as happiness, sadness, anger, fear,

surprise, and neutrality. Understanding these emotions is crucial for intelligent systems interacting with humans.

Traditional emotion recognition systems relied on handcrafted features and classical machine learning algorithms. However, these methods struggled under real-world conditions such as illumination changes, pose variations, and occlusions.

With advancements in **Deep Learning**, especially **Convolutional Neural Networks (CNNs)**, computers can now automatically learn hierarchical facial features directly from images. Facial Emotion Recognition has therefore become more accurate, robust, and scalable.

This project focuses on designing a **deep learning-based FER system** capable of recognizing emotions from static images and real-time webcam video using modern architectures like **ConvNeXt**.

Aim, Objectives, Scope

Aim

To develop a **deep learning-based facial emotion recognition system** capable of accurately classifying human emotions from facial expressions and performing real-time detection using a webcam.

Objectives

- Study existing facial emotion recognition techniques
- Preprocess and analyze the FER2013+ dataset
- Implement CNN and transfer learning models
- Integrate MTCNN for multi-face detection
- Evaluate model performance using standard metrics
- Deploy a real-time emotion recognition application

Scope

- Image and webcam-based facial emotion recognition

- Visual data only (no audio or gesture analysis)
- Eight emotion classes from FER2013+ dataset
- Real-time deployment using Streamlit

System Overview

The Face Emotion Recognition system consists of multiple integrated components working in a sequential pipeline:

Major Components

1. **Data Input Module** – Image upload, video, or webcam input
2. **Face Detection Module** – MTCNN for detecting and aligning faces
3. **Preprocessing Module** – Resizing, normalization, augmentation
4. **Emotion Classification Module** – ConvNeXt deep learning model
5. **Output & Visualization Module** – Emotion labels and confidence scores

System Workflow

1. Capture input image/video
2. Detect faces using MTCNN
3. Crop and preprocess facial regions
4. Extract features using ConvNeXt
5. Predict emotion probabilities
6. Display results in real time

Dataset and Preprocessing

Dataset

- **FER2013+ dataset**
- Image size: **48×48 grayscale**
- Emotion classes: 8
- Split into training and validation sets

Preprocessing Steps

- Duplicate image removal using hashing
- Resizing to model input size
- Grayscale to RGB conversion
- Histogram equalization for contrast enhancement
- Normalization using ImageNet statistics
- Data augmentation (flip, rotation, scaling)

These steps improve model generalization and reduce overfitting.

Face Detection using MTCNN

Face detection is a critical step in emotion recognition, especially for real-time applications.

MTCNN (Multi-task Cascaded CNN)

- Consists of **P-Net, R-Net, and O-Net**
- Detects multiple faces simultaneously

- Performs facial alignment
- Robust under varying lighting and poses

Using MTCNN ensures accurate face localization and improves emotion classification performance.

Model Architecture

Baseline CNN

- Convolution layers
- Batch normalization
- ReLU activation
- Max pooling
- Fully connected layers

Used as a reference model.

ConvNeXt Transfer Learning Model

- Modern CNN inspired by Vision Transformers
- Pretrained on ImageNet
- Strong hierarchical feature extraction
- Fine-tuned for FER2013+ dataset

ConvNeXt significantly outperforms the baseline CNN.

Training Strategy

Training is performed in **two stages**:

Stage 1: Feature Extraction

- Freeze pretrained ConvNeXt layers
- Train classifier layers only

Stage 2: Fine-Tuning

- Unfreeze higher layers gradually
- Lower learning rate for stability

Training Details

- Loss Function: Categorical Cross-Entropy
- Optimizer: AdamW
- Learning Rate Scheduler applied

Fine-tuning improves domain-specific emotion recognition.

Model Evaluation

Evaluation Metrics

- Accuracy
- Precision
- Recall
- F1-score
- Confusion Matrix

Results

- Overall validation accuracy: **~79%**
- High accuracy for **Happy** and **Neutral**
- Lower accuracy for **Contempt** and **Disgust** due to class imbalance

Confusion matrix analysis highlights commonly confused emotions.

Results and Analysis

Model Comparison

Model	Validation Accuracy
Custom CNN	~69%
MobileNetV3	~74%
EfficientNet-B2	~77%
ConvNeXt (Fine-Tuned)	**~79%**

Observations

- Transfer learning improves performance significantly
- Fine-tuning increases generalization
- ConvNeXt offers best balance between accuracy and speed

Features of the System

- Multi-face emotion detection
- Real-time webcam emotion recognition

- Emotion confidence percentage
- Emotion probability visualization
- Support for image and webcam input
- User-friendly Streamlit interface
- Robust preprocessing pipeline

Applications include classroom engagement analysis, healthcare monitoring, and HCI systems.

Conclusion and Future Scope

Conclusion

This project successfully demonstrates a **deep learning-based facial emotion recognition system** using ConvNeXt and MTCNN. The system achieves reliable accuracy and performs effectively in real-time scenarios. It bridges theoretical deep learning concepts with practical deployment.

Future Scope

- Mental health monitoring
- Smart classroom systems
- Emotion-aware human–computer interaction
- Multimodal emotion recognition (face + voice)
- Deployment on mobile and edge devices
- Ethical and bias-aware AI systems