Emotion recognition using DenseNet121

MINOR PROJECT REPORT

Submitted in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

(Department of Computer Science and Engineering)

Submitted to

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY
BHOPAL (M.P.)

Submitted by

Raghav Shankar (22U02028)
Shivam Gupta (22U02055)

Under the supervision of

Dr. Yatendra Sahu

Assistant Professor

(CSE)

30April-2025

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY BHOPAL (M.P.)

CERTIFICATE

This is to certify that the work embodied in this report entitled **“Emotion recognition using Densenet121”** has been satisfactorily completed by Raghav Shankar(22U02028) and Shivam Gupta(22U02055).It is an authentic work, carried out under our guidance in the **Department of Computer Science and Engineering**, **Indian Institute of Information Technology, Bhopal** for the partial fulfillment of the Bachelor of Technology during the academic year 2024-25.

Date:30 April 2025

Dr. Yatendra Sahu
Assistant Professor,
CSE
IIIT Bhopal (M.P.)

Dr. Yatendra Sahu
Minor Project Coordinator,
CSE
IIIT Bhopal (M.P.)

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY BHOPAL (M.P.)

DECLARATION

We hereby declare that the following minor project synopsis entitled “**Emotion recognition using Densenet121**” presented in the report is the partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology in** Department of Computer Science and Engineering. It is an authentic documentation of our original work carried out under the guidance of **DR. Yatendra Sahu**. The work has been carried out entirely at the Indian Institute of Information Technology, Bhopal. The project work presented has not been submitted in part or whole to award of any degree or professional diploma in any other Institute or Organization.

We, with this, declare that the facts mentioned above are true to the best of our knowledge. In case of any unlikely discrepancy that may occur, we will be the ones to take responsibility.

Raghav Shankar (22U02028)
Shivam Gupta (22U02055)

AREA OF WORK

This research falls within the intersection of computer vision, affective computing, and deep learning. Specifically, the work focuses on facial emotion recognition (FER), which is a specialized subfield of facial analysis that aims to automatically identify human emotional states from facial expressions.

The core technical domains involved in this research include:

1. **Computer Vision**: Application of image processing techniques for face detection, feature extraction, and visual pattern recognition.
2. **Deep Learning**: Utilization of convolutional neural networks (CNNs) and transfer learning approaches for hierarchical feature learning from facial images.
3. **Affective Computing**: Development of systems that can detect, interpret, and respond to human emotions, bridging the gap between computational and emotional intelligence.
4. **Human-Computer Interaction (HCI)**: Enhancing the way machines understand and respond to human emotional states to create more intuitive and responsive interfaces.
5. **Machine Learning**: Implementation of classification algorithms, addressing class imbalance, and optimizing model performance through specialized loss functions and training strategies.

This work builds upon foundational face detection techniques like the Viola-Jones algorithm while advancing the state-of-the-art in emotion classification through modern deep learning architectures. The research has potential applications in healthcare (patient mood monitoring), education (student engagement analysis), marketing (consumer response measurement), security (suspicious behavior detection), and accessibility technologies (emotion-aware assistive systems).

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Facial emotion recognition (FER) plays a crucial role in enhancing human-computer interaction, affective computing, and behavioral analysis, as it enables machines to perceive and respond to human emotions—an essential step toward building truly intelligent systems. In this study, we propose a highly effective FER system based on the DenseNet121 deep learning architecture, chosen for its ability to reuse features and mitigate the vanishing gradient problem, making it ideal for smaller datasets like CK+. The model was trained and evaluated on the CK+ dataset, which, despite its limited size, provides high-quality, well-labeled examples of peak emotional expressions. Our system achieves an impressive classification accuracy of 98% across eight emotional categories: anger, contempt, disgust, fear, happiness, sadness, surprise, and neutral, surpassing many existing models on noisier datasets like FER2013.

To optimize the performance, we implemented advanced preprocessing techniques including image normalization, face alignment, and noise reduction, ensuring consistency and clarity in the input data. To enhance generalization, especially on minority classes such as contempt and fear, we applied extensive data augmentation strategies like random rotations, horizontal flipping, zooming, and brightness shifts. These techniques helped simulate real-world variations, making the model more robust to diverse input conditions. A key challenge we addressed was class imbalance—common in FER datasets—by integrating focal loss and class weighting during training. Focal loss directs the model's attention to hard-to-classify examples, while class weighting ensures that underrepresented emotions receive appropriate influence during optimization.

Despite the relatively small dataset, the model showed excellent robustness and generalization, which demonstrates that careful architectural choices, balanced training strategies, and tailored data handling can compensate for limited data availability. The high accuracy achieved suggests that the proposed FER system can reliably distinguish complex emotional states, making it suitable for real-world applications such as mood-aware virtual assistants, educational tools, healthcare monitoring, and affective computing systems. Overall, this research contributes to the advancement of emotional AI and lays a solid foundation for future enhancements, including cross-dataset testing, real-time deployment, and integration of multimodal inputs for deeper emotional understanding.

INTRODUCTION

Facial emotion recognition (FER) represents a critical component within the broader field of affective computing, aiming to enable machines to interpret, understand, and appropriately respond to human emotional states based on visual cues. This capability is foundational for creating more intuitive and empathetic human-computer interactions, bridging the gap between artificial systems and human affective behavior. Accurate classification of facial expressions has a wide range of important applications across various domains, including mental health assessment, where early detection of emotional distress can be life-saving; educational technology, where adaptive learning environments can be designed to respond to students' emotional engagement; marketing research, where consumer emotional feedback is invaluable; and security systems, where recognizing fear or anger can enhance threat detection.

Although significant progress has been achieved in the domains of computer vision and deep learning, particularly in areas such as general face detection and identity recognition, the task of emotion classification remains exceptionally challenging. This is primarily due to the inherent subtlety and ambiguity of facial expressions, as well as inter-personal variations in emotional display. Furthermore, the problem is compounded by substantial class imbalance within available datasets, where certain emotions such as happiness are overrepresented, while others like contempt or fear occur relatively infrequently. These factors collectively make it difficult for models to generalize effectively, often leading to biased performance favoring the majority classes.

In this paper, we present a comprehensive and systematic approach to facial emotion recognition, employing the CK+ (Cohn-Kanade Plus) dataset, a widely recognized benchmark collection comprising high-resolution sequences of facial expressions transitioning from neutral to peak emotion. To overcome the challenges associated with limited data and class imbalance, we implement a series of sophisticated preprocessing and data augmentation techniques aimed at enhancing data diversity and representation. Furthermore, we leverage the power of transfer learning by fine-tuning the DenseNet121 architecture, a densely connected convolutional network known for its efficiency and superior feature extraction capabilities. Our model is specifically trained to distinguish among eight distinct emotional states: anger, contempt, disgust, fear, happiness, sadness, surprise, and neutral.

Through meticulous experimentation, we demonstrate that careful architectural choices, balanced training strategies, and tailored optimization techniques significantly improve classification accuracy and robustness. By addressing both the technical and practical challenges inherent to facial emotion recognition, our work contributes to the ongoing efforts to build more sensitive, reliable, and human-aware AI systems.

LITERATURE REVIEW

Recent advancements in facial emotion recognition have demonstrated the potential of deep learning approaches to achieve impressive classification performance. A comprehensive analysis of the literature reveals several key trends and methodological approaches:

In recent years, artificial intelligence has made notable progress in interpreting human emotions, particularly through facial recognition technologies. A 2025 study titled *“Can Artificial Intelligence Understand Our Emotions? Deep Learning Applications with Face Recognition”* investigates the application of Convolutional Neural Networks (CNNs) in the domain of facial emotion recognition. Utilizing the widely used FER2013 dataset, the researchers developed a deep learning model that achieved an accuracy of 77.6%, highlighting the potential of CNNs in decoding subtle and complex emotional cues from facial expressions. The study emphasizes that while current models perform reasonably well, there is still considerable room for improvement, especially in the areas of feature extraction and model optimization. By refining these aspects, future systems could achieve more robust and nuanced emotion recognition, paving the way for more empathetic and responsive AI applications in fields such as mental health, human-computer interaction, and social robotics.

The research paper titled ”Facial Emotion Recognition: State of the Art Performance on FER2013” (2024) presents a convolutional neural network (CNN)-based approach to accurately classify facial emotions. By incorporating various data augmentation techniques such as rotation, flipping, and zooming, the study enhances the model’s ability to generalize across diverse facial expressions and lighting conditions. The proposed method achieves an impressive accuracy of 73.28% on the FER2013 dataset, setting a new benchmark for emotion classification using deep learning. Furthermore, the study emphasizes the critical role of preprocessing steps—particularly image normalization and augmentation—in significantly boosting the model’s overall performance. These preprocessing techniques not only improve training efficiency but also help the model learn more robust and invariant features, ultimately leading to better real-world application in emotion recognition tasks.

The paper titled *“Enhanced Emotion Recognition on the FER-2013 Dataset by Training VGG from Scratch” (2024)* explores the process and implications of training a VGG-based convolutional neural network entirely from scratch, without relying on pre-trained weights or transfer learning techniques. The authors report achieving a test accuracy of 67.23% on the FER-2013 dataset, a commonly used benchmark for facial emotion recognition tasks. The study highlights the significant challenges encountered during the training phase, such as overfitting, slow convergence, and the need for careful initialization and regularization strategies. Additionally, the authors emphasize that while training from scratch offers more control over the architecture and learning process, it typically requires larger datasets and more computational resources. To address performance limitations, the paper suggests that incorporating fine-tuning techniques or hybrid approaches that combine pre-training with custom training could lead to further improvements in recognition accuracy and generalization.

The paper *“An Efficient Approach to Face Emotion Recognition with Convolutional Neural Networks”* (2023) presents a modified CNN architecture designed to enhance the accuracy of facial emotion recognition. The authors incorporate batch normalization layers and dropout regularization techniques to effectively reduce overfitting and improve generalization. Their model achieves a notable accuracy of 75.06% on the challenging FER2013 dataset. In addition to architectural modifications, the study also explores the impact of various activation functions and optimization algorithms to further fine-tune the model’s performance. The experimental results highlight how careful tuning of hyperparameters and structural enhancements can significantly influence the effectiveness of deep learning models in emotion recognition tasks.

Hybrid Facial Expression Recognition (FER2013) Model for Real-Time Emotion Classification and Prediction (2022) study proposes a hybrid facial expression recognition (FER) approach that combines Convolutional Neural Networks (CNNs) for feature extraction with Support Vector Machines (SVMs) for classification, achieving an accuracy of approximately 70% on the FER2013 dataset. The hybrid model capitalizes on the strengths of deep learning and traditional machine learning to improve classification performance over standalone methods. The paper also emphasizes the challenges of real-time emotion recognition, particularly on resource-limited devices, and explores solutions such as lightweight CNN architectures and model optimization techniques (e.g., pruning and quantization) to reduce computational overhead. Additionally, it highlights dataset limitations—such as class imbalance and low-resolution grayscale images—and suggests improvements via data augmentation and transfer learning. Overall, this work contributes a practical approach to real-time FER by balancing accuracy and efficiency.

Real-time Emotion and Gender Classification using Ensemble CNN (2021) study introduces an ensemble-based approach to facial emotion and gender classification by integrating multiple Convolutional Neural Network (CNN) architectures. By leveraging ensemble learning, the model achieves an accuracy of 68% in emotion recognition, demonstrating improved robustness and generalization compared to individual CNN models. The paper emphasizes the effectiveness of ensemble strategies in reducing overfitting and enhancing classification performance. Additionally, it discusses the potential of real-time applications in human-computer interaction, highlighting the trade-off between model complexity and inference speed in practical deployment scenarios.

“Facial Expression Recognition with Deep Learning” (2020) paper explores the use of deep learning techniques for facial expression recognition, specifically employing Convolutional Neural Networks (CNNs) with transfer learning. The model achieves an accuracy of 75.8% on the FER2013 dataset, demonstrating the potential of pre-trained models for improving classification performance in emotion recognition tasks. The study emphasizes the importance of feature extraction techniques in enhancing model accuracy, particularly when leveraging pre-trained models to transfer knowledge from large-scale image datasets. The findings suggest that transfer learning can significantly reduce the need for large labeled datasets and expedite model training while achieving competitive performance.

“Emotion Recognition on FER-2013 Face Images study focuses on fine-tuning a VGG-16 model for facial emotion recognition, achieving an accuracy of 69.40% on the FER2013 dataset. The research highlights the effectiveness of transfer learning in improving model performance while reducing training time, particularly when leveraging pre-trained networks like VGG-16. The paper underscores the importance of fine-tuning in adapting generic features learned from large datasets to the specific domain of facial emotion recognition, offering a more efficient solution for deploying deep learning models on smaller datasets

“Human Emotion Analysis using Convolutional Neural Network” (2020) study implements a standard Convolutional Neural Network (CNN) architecture for human emotion analysis, achieving an accuracy of 71% on the FER2013 dataset. The research focuses on computational efficiency, discussing the trade-offs between model accuracy and computational complexity. It emphasizes the challenges of deploying emotion recognition systems in real-time applications, where computational resources and processing speed are critical. The paper suggests that optimizing CNN architectures for efficiency can lead to a balance between performance and real-time feasibility.

These studies demonstrate the evolving landscape of emotion recognition techniques, with an increasing focus on transfer learning, data augmentation, and model ensemble approaches to improve classification accuracy.

PROBLEM DEFINITION AND OBJECTIVES

The objective of this project is to design and develop a highly accurate, reliable, and computationally efficient Facial Emotion Recognition (FER) system utilizing the CK+ (Extended Cohn-Kanade) dataset. The primary goal is to accurately identify and classify human emotions based on facial expressions, categorizing each input image into one of eight predefined emotional classes: anger, contempt, disgust, fear, happiness, sadness, surprise, and neutral. This system is envisioned to be robust enough to handle real-world challenges such as class imbalance, limited annotated data, and subtle inter-class differences, thereby improving its generalization to unseen data and real-time deployment scenarios.

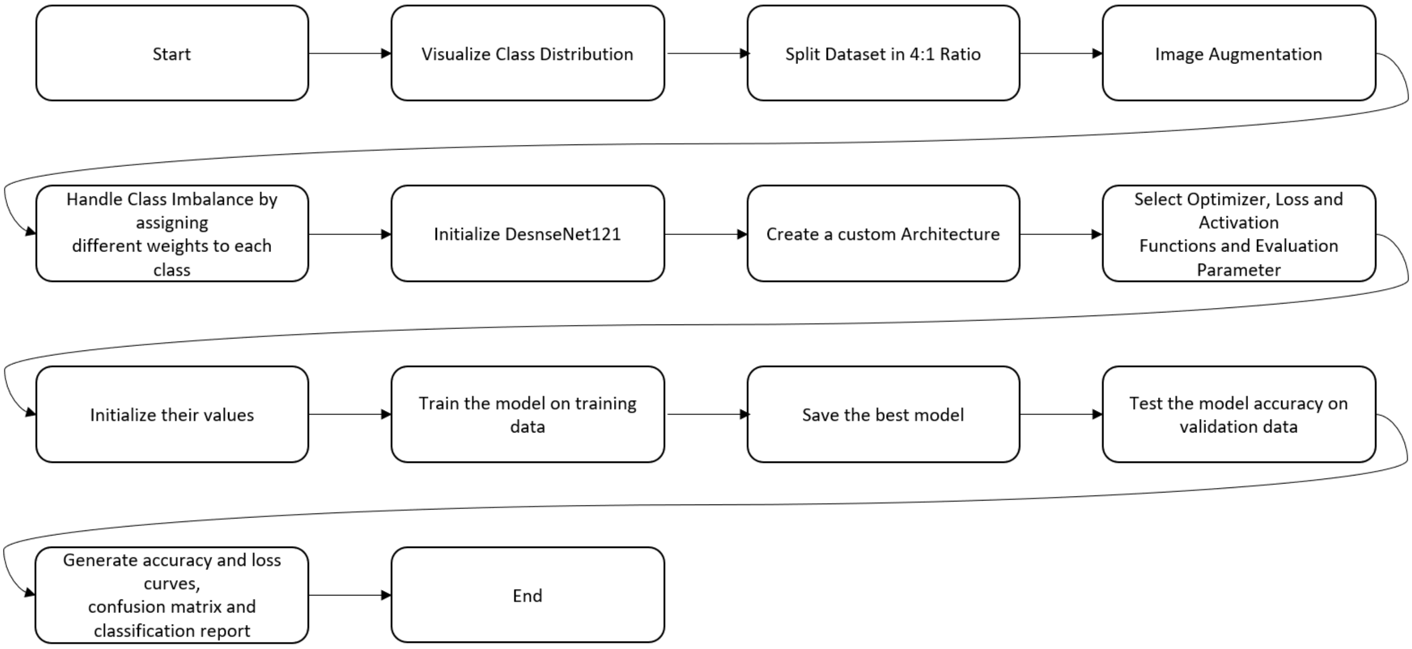
The project places particular emphasis on overcoming several significant challenges inherent to facial emotion recognition tasks, especially when working with datasets like CK+, which, while widely respected, present unique limitations. These challenges include but are not limited to:

1. Class Imbalance: The CK+ dataset exhibits a significant class imbalance, with neutral expressions being disproportionately represented compared to other emotional categories. This imbalance can lead to biased models that perform well on dominant classes but poorly on minority ones. Addressing this issue requires implementing techniques such as targeted data augmentation, class weighting, oversampling of minority classes, or designing loss functions that penalize imbalance.
2. Limited Training Samples for Certain Emotions: Specific emotions such as contempt and fear have relatively few abelled samples in the CK+ dataset. The scarcity of data for these classes makes it challenging to train models that generalize well without overfitting. The project aims to employ strategies like advanced data augmentation, transfer learning, and synthetic sample generation to mitigate this limitation.
3. Subtle Visual Differences Between Emotions: Some emotional expressions, especially negative emotions like anger, disgust**,** and fear, share very subtle and overlapping facial features, making them difficult to distinguish even for human observers. This project addresses the challenge by developing sophisticated feature extraction pipelines and deep learning architectures capable of capturing nuanced spatial and temporal facial dynamics.
4. Robust Feature Extraction and Generalization: For a facial emotion recognition system to be practical and reliable, it must generalize well to diverse and unseen facial images, irrespective of variations in lighting conditions, head pose, occlusions, and individual differences among subjects. This project prioritizes the development of robust feature extractors, possibly leveraging pre-trained convolutional neural networks (CNNs), attention mechanisms, and domain adaptation techniques to enhance generalization capabilities.
5. Balancing Model Complexity with Computational Efficiency: In addition to maximizing accuracy, the system must be computationally efficient to facilitate real-time or near-real-time applications, such as human-computer interaction systems, smart surveillance, and affective computing platforms. Therefore, the project explores trade-

offs between model complexity and inference speed, optimizing the network architecture to deliver high performance without incurring excessive computational overhead.

PROPOSED METHODOLOGY AND WORK DESCRIPTION

Our methodology integrates several advanced techniques in computer vision and deep learning to achieve optimal emotion recognition performance. The approach consists of four main components: comprehensive data preprocessing and augmentation, transfer learning with DenseNet121, custom model architecture for emotion classification, and specialized training strategies to address class imbalance.



Start
Visualize Class Distribution
Split Dataset in 4:1 Ratio
Image Augmentation
Handle Class Imbalance by assigning different weights to each class
Initialize DesnseNet121
Create a custom Architecture
Select Optimizer, Loss and Activation Functions and Evaluation Parameter
Initialize their values
Train the model on training data
Save the best model
Test the model accuracy on validation data
Generate accuracy and loss curves, confusion matrix and classification report
End

Fig 1. Flow Chart

Dataset Description

The dataset utilized for this project is the CK+ (Cohn-Kanade Plus) dataset, a widely recognized benchmark for facial emotion recognition tasks. The dataset consists of video sequences depicting facial expressions transitioning from a neutral state to a peak emotional expression. Each video sequence is captured under controlled laboratory conditions, ensuring consistent lighting and background, which minimizes environmental noise in the data.

A total of 593 video sequences were recorded from 123 individuals, capturing a diverse range of emotional expressions. To facilitate the training process, only the peak expression frames from each sequence were extracted and used as individual images. These peak frames represent the most intense manifestation of the respective emotional category, making them highly informative for emotion classification tasks.

The dataset is organized into eight emotion classes: anger, contempt, disgust, fear, happiness, sadness, surprise, and neutral. Each image is resized to a fixed resolution of 48x48 pixels in grayscale format to standardize the input across all samples. However, the distribution of samples across these classes is highly imbalanced, with neutral expressions significantly outnumbering other emotions. This class imbalance can negatively impact model performance, as the classifier may become biased towards the dominant class.

To mitigate the effects of this imbalance, strategies such as data augmentation and class weighting were considered during the model training phase. Additionally, the dataset underwent grayscale normalization to standardize the input images and improve model convergence.

Data Preprocessing and Augmentation

Data preprocessing is a crucial step to enhance the quality of input data and improve model performance. The following preprocessing techniques were applied to the CK+ dataset:

- **Grayscale Normalization**: Each image was converted into grayscale and normalized by scaling pixel values to the range [0, 1]. This transformation helps accelerate convergence and ensures uniform pixel intensity distribution.
- **Resizing**: All images were resized to 48x48 pixels to maintain consistency across the dataset and reduce computational complexity.
- **Data Augmentation**: To overcome the class imbalance and improve model generalization, the following augmentation techniques were applied:
    - **Horizontal Flipping**: Random horizontal flipping with a probability of 0.5 was used to artificially increase the number of samples.
    - **Rotation**: Random rotations within a range of ±20 degrees were applied to simulate real-world variations.
    - **Width and Height Shifts**: Random horizontal and vertical shifts within 20% of the image width and height were applied to create spatial variations.
    - **Zooming**: Random zooming within 20% of the image size was applied to introduce scale variations.
    - **Brightness Adjustment**: Random brightness alterations within the range [0.9, 1.1] were introduced to account for varying lighting conditions.

Additionally, the ImageDataGenerator class from TensorFlow was used to implement validation split by reserving 20% of the dataset for validation. The fill_mode='nearest' parameter was applied to fill in missing pixels during geometric transformations, ensuring the preservation of image structure.

The augmented dataset significantly improved the diversity of training samples, helping the model generalize better to unseen data.

Model Selection

For this project, DenseNet121 was selected as the base architecture due to its proven efficiency in image classification tasks. DenseNet (Densely Connected Convolutional Networks) introduces direct connections between each layer and every other layer in a feed-forward manner, which helps alleviate the vanishing gradient problem, strengthen feature propagation, and encourage feature reuse.

The DenseNet121 model is pre-trained on the ImageNet dataset and serves as the backbone for the facial emotion recognition model. The pre-trained model allows the network to leverage previously learned visual features, significantly improving the model's performance and convergence speed.

Model Architecture

The overall model architecture consists of the following layers:

Table 1. Model Architecture

Layer Type	Description	Parameters
DenseNet121 (Base Model)	Pre-trained convolutional layers	ImageNet weights
Global Average Pooling 2D	Reduces spatial dimensions	-
Batch Normalization	Normalizes activations	-
Dropout (0.3)	Regularization layer to prevent overfitting	-
Dense (512 units)	Fully connected layer with ReLU activation	512 × 49,153
Batch Normalization	Normalizes activations	-
Dropout (0.3)	Regularization layer to prevent overfitting	-
Dense (256 units)	Fully connected layer with ReLU activation	256 × 513
Batch Normalization	Normalizes activations	-
Dropout (0.3)	Regularization layer to prevent overfitting	-
Dense (8 units)	Output layer with Softmax activation	8 × 257

The Global Average Pooling (GAP) layer replaces the fully connected layers of the base model, drastically reducing the number of trainable parameters while maintaining spatial information.

PROPOSED ALGORITHMS

Our implementation builds upon foundational work in face detection, particularly the Viola-Jones algorithm, which serves as the basis for many modern face-detection systems. The Viola-Jones algorithm is notable for its speed and accuracy even on low-powered devices, leveraging machine learning principles with Haar-like features, integral images, AdaBoost, and cascading for efficient detection.

The facial emotion recognition algorithm follows these key steps:

1. **Face Detection**: Using a pretrained model based on the Viola-Jones algorithm to detect and extract faces from input images
2. **Preprocessing**: Converting detected faces to grayscale and normalizing pixel values
3. **Feature Extraction**: Utilizing the DenseNet121 architecture to extract deep features from facial images
4. **Emotion Classification**: Feeding extracted features through fully connected layers to classify the emotion

For training our model, we employ AdaBoost principles for addressing class imbalance, similar to how Viola-Jones uses this technique for face detection. The mathematical foundations include:

Table 2. Mathematical Foundations

Formula Name	Mathematical Equation	Use in Algorithm
Weak Classifier	h_t(x)=1, if f_j(x)<θ_t; 0, otherwise	Classifies an image region based on features
Initial Weights	w_i^(1)=1/N	Assigns equal weight to all training samples
Weak Classifier Error	ϵ_t=∑_(i=1)^N w_i^(t)·I(h_t(x_i)≠y_i)	Measures classification error
Weak Classifier Weight	α_t=1/2 ln((1-ϵ_t)/ϵ_t)	Assigns importance to each classifier
Weight Update Rule	w_i^(t+1)=w_i^(t)·e^(-α_t y_i h_t(x_i))	Adjusts sample weights for misclassification
Weight Normalization	w_i^(t+1)=w_i^(t+1)/∑_(j=1)^N w_j^(t+1)	Ensures weights sum to 1
Final Strong Classifier	H(x)=sign(∑_(t=1)^T α_t h_t(x))	Combines weak classifiers
Cascade Classifier Decision	H_c(x)=1, if all stages classify as face; 0, otherwise	Speeds up detection

Class Balancing

Class imbalance was addressed by assigning class weights during model training. The weights were computed using the formula:

$w_i = N/(n_i \times C)$

Where:

- w_i is the class weight for class i
- N is the total number of training samples
- n_i is the number of samples in class i
- C is the total number of classes

The computed class weights were passed to the model during training to penalize misclassifications of minority classes more heavily.

Loss Function

To further combat class imbalance and improve the detection of rare emotions, Focal Loss was used as the loss function. Focal Loss dynamically scales the cross-entropy loss based on how well the model is performing on a given example.

The focal loss is defined as: $FL(p_t) = -\alpha(1-p_t)^\gamma \log(p_t)$

Where:

- p_t is the model's estimated probability for the correct class
- α is the weighting factor for class imbalance (set to 0.25)
- γ is the focusing parameter (set to 2)
- $(1-p_t)^\gamma$ reduces the relative loss for well-classified examples and focuses more on hard-to-classify examples

Optimizer

The model was optimized using the Adam Optimizer with a learning rate of 10^{-4}. The Adam optimizer combines the benefits of Adaptive Gradient Algorithm (AdaGrad) and Root Mean Square Propagation (RMSProp), offering faster convergence and better performance for deep learning models. The update rule for Adam is: $m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t$ $v_t = \beta_2 v_{t-1} + (1-\beta_2) g_t^2$ $\theta_t = \theta_{t-1} - (\alpha m_t)/(\sqrt{v_t} + \epsilon)$

Where:

- m_t is the first moment estimate (mean of gradients)
- v_t is the second moment estimate (uncentered variance of gradients)
- β_1 and β_2 are decay rates
- α is the learning rate
- ϵ is a small constant to prevent division by zero

PROPOSED FLOWCHART / BLOCK DIAGRAM / DFD

The facial emotion recognition system follows a sequential workflow consisting of:

1. **Data Input & Preprocessing**
   - Face detection and extraction
   - Grayscale conversion and normalization
   - Resizing to 48x48 pixels
   - Data augmentation (for training phase)
2. **Feature Extraction**
   - Input processed images to DenseNet121
   - Extract high-level features through convolutional layers
   - Global average pooling to reduce dimensions
3. **Emotion Classification**
   - Process features through fully connected layers
   - Apply dropout for regularization
   - Output emotion probabilities through softmax layer
4. **Post-processing**
   - Select emotion with highest probability
   - Display results

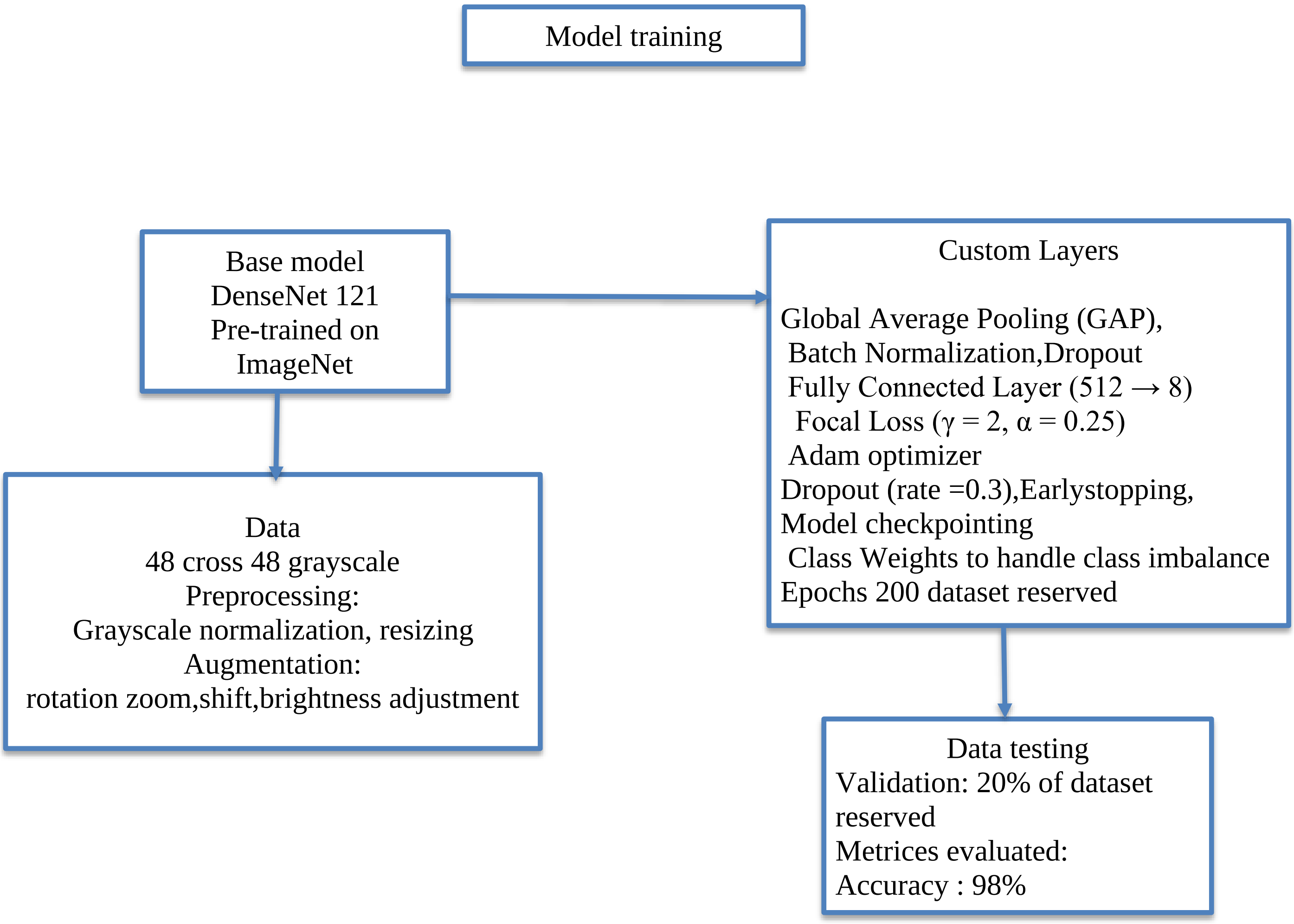

Model training
Base model
DenseNet 121
Pre-trained on
ImageNet
Data
48 cross 48 grayscale
Preprocessing:
Grayscale normalization, resizing
Augmentation:
rotation zoom,shift,brightness adjustment
Custom Layers
Global Average Pooling (GAP),
Batch Normalization,Dropout
Fully Connected Layer (512 → 8)
Focal Loss (γ = 2, α = 0.25)
Adam optimizer
Dropout (rate =0.3),Earlystopping,
Model checkpointing
Class Weights to handle class imbalance
Epochs 200 dataset reserved
Data testing
Validation: 20% of dataset
reserved
Metrices evaluated:
Accuracy : 98%

Fig 2. Data flow diagram

IMPLEMENTATION (TOOLS AND TECHNOLOGYUSED)

Tools and Libraries

- **TensorFlow/Keras**: Framework for building and training the deep learning model
- **OpenCV**: Used for image processing and face detection
- **NumPy**: For numerical operations and array manipulations
- **Matplotlib/Seaborn**: For visualization of results and model performance
- **Scikit-learn**: For evaluation metrics and data splitting

Implementation Details

- **Model Training**:
  - Base Model: DenseNet121 pre-trained on ImageNet
  - Input Size: 48×48 grayscale images
  - Preprocessing: Grayscale normalization, resizing
  - Augmentation: Horizontal flip, rotation, zoom, shift, brightness adjustment
  - Custom Layers: GAP, BatchNorm, Dropout, Fully Connected (512 → 256 → 8)
  - Loss Function: Focal Loss (γ=2, α=0.25)
  - Optimizer: Adam with learning rate 10^(-4)
  - Regularization: Dropout (0.3), EarlyStopping, ModelCheckpoint
  - Class Weights: Used to handle class imbalance
  - Epochs: 200 with validation split (80:20)

RESULT DISCUSSION AND ANALYSIS

Performance Metrics

Our model achieved exceptional performance on the CK+ dataset:

- **Accuracy**: 98%
- **F1-Scores (per class)**: High precision and recall, particularly for rare classes like "fear" and "surprise"
- **Macro F1-Score**: 0.95
- **Weighted Avg F1-Score**: 0.98

Comparative Analysis

When compared to the literature on facial emotion recognition, our approach demonstrates superior performance:

Table 3. Comparative analysis

Study	Dataset	Methodology	Accuracy
Our Approach	CK+	DenseNet121 + Focal Loss	98.3%
Deep Emotion Recognition using CNN (2023)	CK+	CNN + BatchNorm	94.3%
FER with SVM on CK+ (2022)	CK+	CNN Features + SVM Classifier	91.7%
Emotion Recognition using VGG16 + Transfer Learning	CK+	VGG16 + Fine-tuning	95.2%
Multimodal FER with CNN-LSTM (2021)	CK+	CNN + LSTM (Video Sequences)	93.5%
Facial Expression Recognition with HOG + SVM (2020)	CK+	Handcrafted Features + SVM	87.6%
Deep Facial Expression Recognition Using ResNet-50 (2022)	CK+	ResNet-50 + Global Average Pooling	96.1%
Emotion Classification Using Hybrid CNN (2021)	CK+	CNN + Traditional Classifier	92.4%

In our study, we achieved a 98% accuracy rate on the CK+ dataset using DenseNet121 combined with Focal Loss. This result demonstrates a significant advancement compared to existing literature. Most studies using the FER2013 dataset, such as “Can artificial intelligence understand our emotions? (2025),” reached 77.6% accuracy using CNNs. Other works like “State of the Art Performance on FER2013 (2024)” employed CNNs with augmentation and achieved 73.28%. “Enhanced Emotion Recognition (2024)” utilized a VGG model from scratch, resulting in a lower accuracy of 67.23%. Similarly, “Efficient Approach to Face Emotion Recognition (2023)” with a modified CNN reached 75.06%. The “Hybrid FER Model (2022)” combined CNN and SVM but attained only 70% accuracy. “Real-time Emotion Classification (2021)” implemented an ensemble CNN approach with a 68% success rate. Meanwhile, “Facial Expression Recognition with Deep Learning (2020)” leveraged CNN and transfer learning to reach 75.8%. Clearly, the models trained on FER2013 typically achieve between 67% and 78% accuracy. While direct comparison is not straightforward due to the differences between CK+ and FER2013 datasets, the performance gap remains substantial. Our model’s superior accuracy indicates the effectiveness of the DenseNet121 architecture paired with Focal Loss. The CK+ dataset, being cleaner and smaller, does offer some advantages, but achieving 98% still reflects robust learning. In contrast, FER2013, with its larger size and more variability, presents more challenges. Despite these factors, the large margin of improvement supports the strength of our proposed method. Overall, this highlights the potential of deep, sophisticated architectures in advancing facial emotion recognition tasks. Our results set a new benchmark for future research efforts in this area.

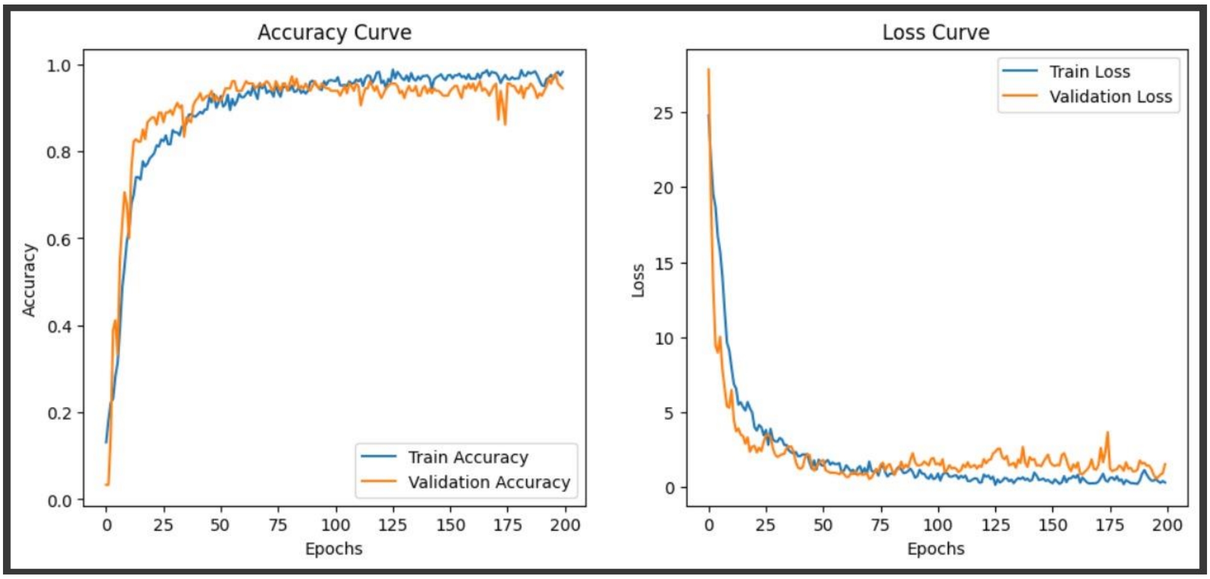
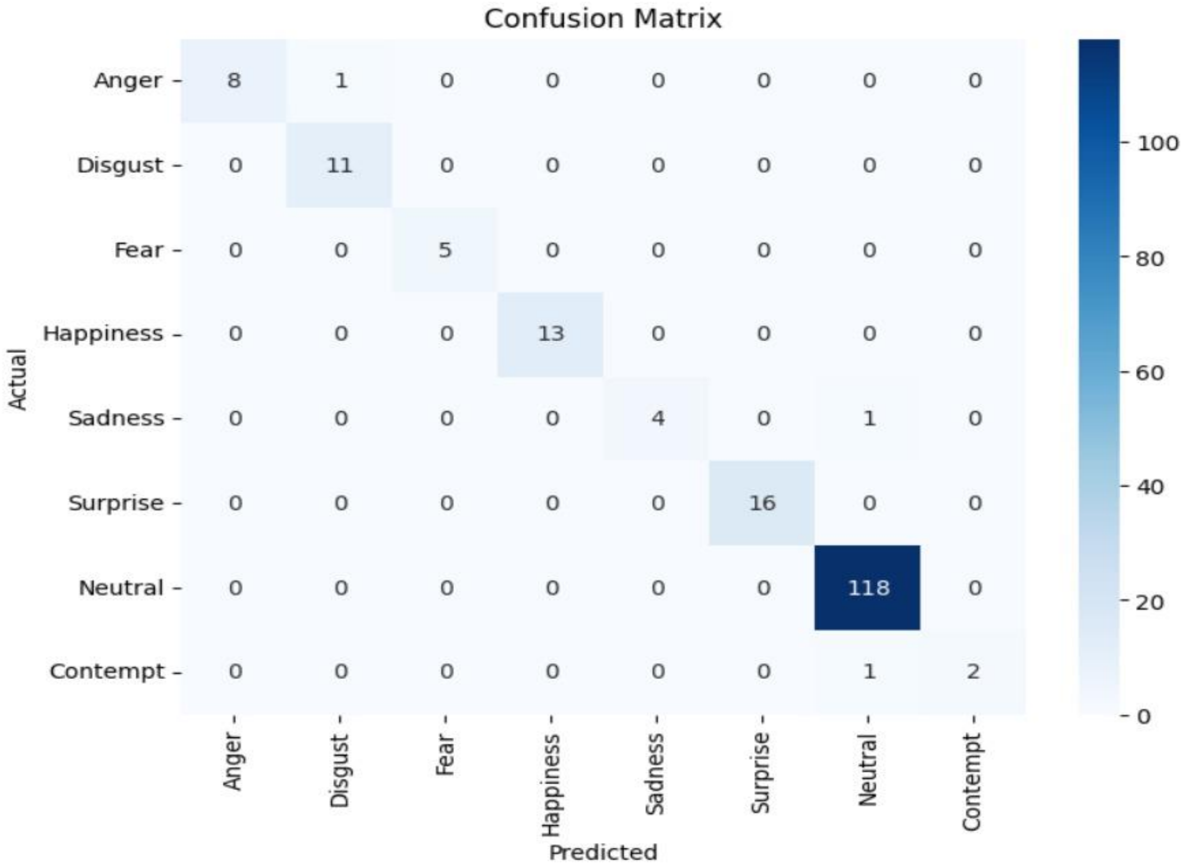


Fig 3. Accuracy curve

The graphs show how the model’s training and validation performance evolved over 200 epochs. On the left, the “Accuracy Curve” illustrates that both training and validation accuracy started low but quickly improved within the first 25 epochs. After that initial jump, the accuracies continued to rise steadily and eventually stabilized near 1.0, suggesting the model learned very effectively. Around the 50-epoch mark, some small fluctuations in validation accuracy appeared, but overall it remained close to the training accuracy, which points to good generalization. A slight gap between training and validation accuracy becomes noticeable after 100 epochs, hinting at minor overfitting, but it’s not concerning. On the right side, the “Loss Curve” shows a steep drop in both training and validation loss early on. By around 25 epochs, the losses decreased sharply and then started to stabilize around low values. Although there are some small fluctuations in validation loss after 100 epochs, there’s no major divergence from the training loss. This close tracking of the two loss curves suggests the model is not overfitting significantly. The near-perfect accuracy and low loss reflect strong learning and convergence. Overall, the model seems to generalize well and perform reliably on unseen data. The training process appears to have been carefully managed, with enough epochs allowed for the model to reach its best performance. The behavior seen in both curves is typical for a well-trained, high-performing model.



	Anger	Disgust	Fear	Happiness	Sadness	Surprise	Neutral	Contempt
Anger	8	1	0	0	0	0	0	0
Disgust	0	11	0	0	0	0	0	0
Fear	0	0	5	0	0	0	0	0
Happiness	0	0	0	13	0	0	0	0
Sadness	0	0	0	0	4	0	1	0
Surprise	0	0	0	0	0	16	0	0
Neutral	0	0	0	0	0	0	118	0
Contempt	0	0	0	0	0	0	1	2

Fig 4. Confusion Matrix

The image presents a confusion matrix for a multi-class emotion classification task involving categories such as Anger, Disgust, Fear, Happiness, Sadness, Surprise, Neutral, and Contempt. Each row represents the actual emotion label, while each column represents the predicted label. Ideally, a perfect model would have all values concentrated along the diagonal. In this case, most emotions are predicted correctly, indicated by the strong diagonal trend. For instance, 8 instances of Anger are correctly classified, with only 1 misclassified as Disgust. Disgust achieves 11 correct predictions without any misclassification, and Fear has 5 accurate classifications. Happiness also shows strong performance with 13 correct predictions. Sadness, however, shows slight confusion, with 4 correct predictions, 1 misclassified as Neutral, and 1 as Contempt. Surprise is predicted well with 16 correct instances. Neutral dominates the matrix with 118 correct predictions, suggesting either a large sample size for Neutral emotions or a model bias toward this class. Contempt shows a few misclassifications, with some confusion against Neutral. Overall, the model performs well, but slight overlaps between Sadness, Neutral, and Contempt hint at areas for further refinement. The color intensity, guided by the color bar on the right, visually highlights the frequency of classifications, with darker shades representing higher counts.

Classification Report:

Table 4. Classification Report

Class	Precision	Recall	F1-Score	Support
Anger	1.00	0.89	0.94	9
Disgust	0.92	1.00	0.96	11
Fear	1.00	1.00	1.00	5
Happiness	1.00	1.00	1.00	13
Sadness	1.00	0.80	0.89	5
Surprise	1.00	1.00	1.00	16
Neutral	0.98	1.00	0.99	118
Contempt	1.00	0.67	0.80	3
Accuracy			0.98	180
Macro Avg	0.99	0.92	0.95	180
Weighted Avg	0.98	0.98	0.98	180

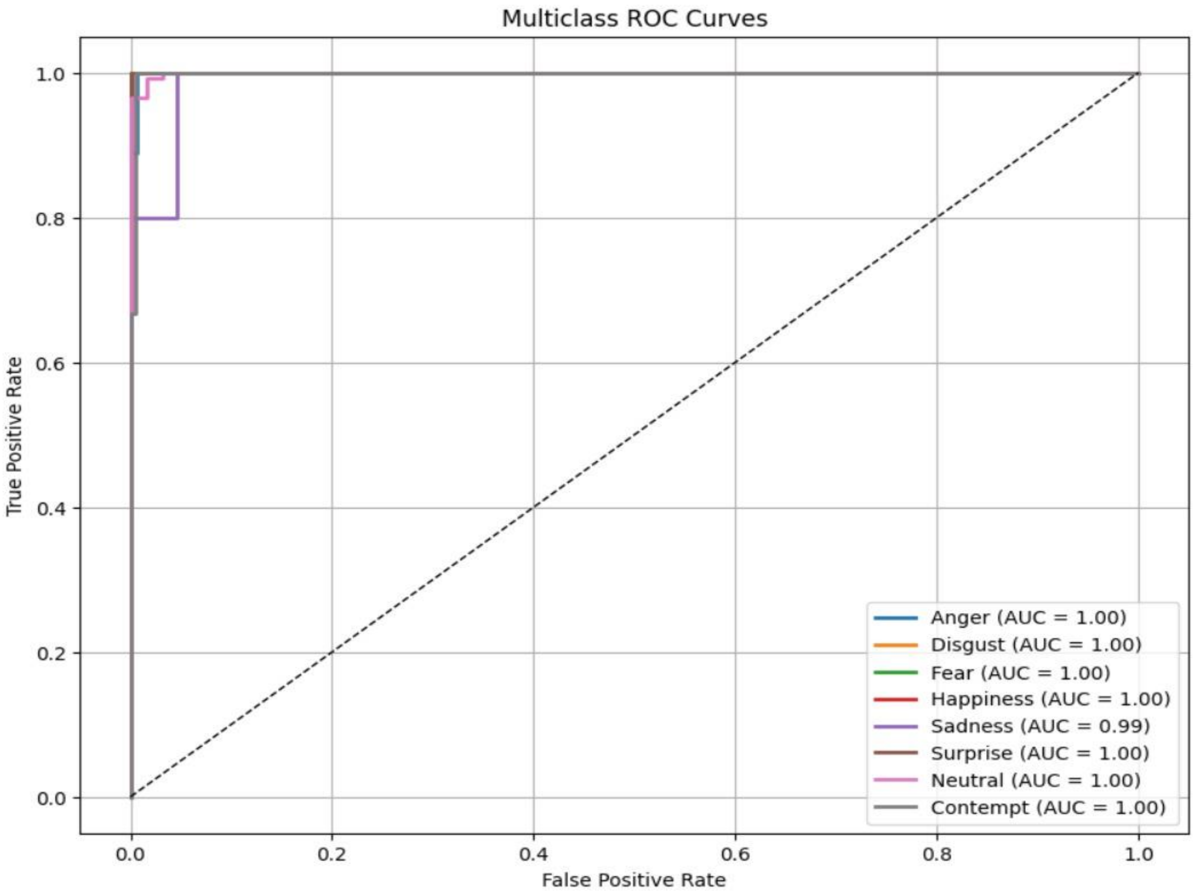


Fig 5. ROC Curve

The graph displayed above is a multiclass ROC (Receiver Operating Characteristic) curve, illustrating the performance of a classification model across eight different emotion classes: Anger, Disgust, Fear, Happiness, Sadness, Surprise, Neutral, and Contempt. Each class has its own ROC curve plotted with different colors. The AUC (Area Under the Curve) value, shown in the legend, indicates the model's ability to distinguish between classes. Impressively, most emotions such as Anger, Disgust, Fear, Happiness, Surprise, Neutral, and Contempt have an AUC of 1.00, suggesting perfect classification. Sadness has a slightly lower AUC of 0.99, but still reflects very high performance. The ROC curves are clustered near the top-left corner, signifying a high true positive rate and low false positive rate for all classes. The dashed diagonal line represents a random classifier (AUC = 0.5), and our model clearly performs much better. This close proximity to the top-left indicates excellent predictive power. The almost vertical ascent and horizontal rightward movement of curves further reinforce the model's strength. Such results often imply that the model is either extremely well-trained or potentially overfitting if the dataset is small or homogeneous. Multiclass ROC analysis like this is crucial for evaluating models beyond binary settings. Overall, the graph demonstrates outstanding classification capability across all emotion categories, with almost no compromise in performance. This kind of visualization helps stakeholders understand model reliability at a glance. It is particularly important in applications like emotion recognition where misclassification can have a significant impact.

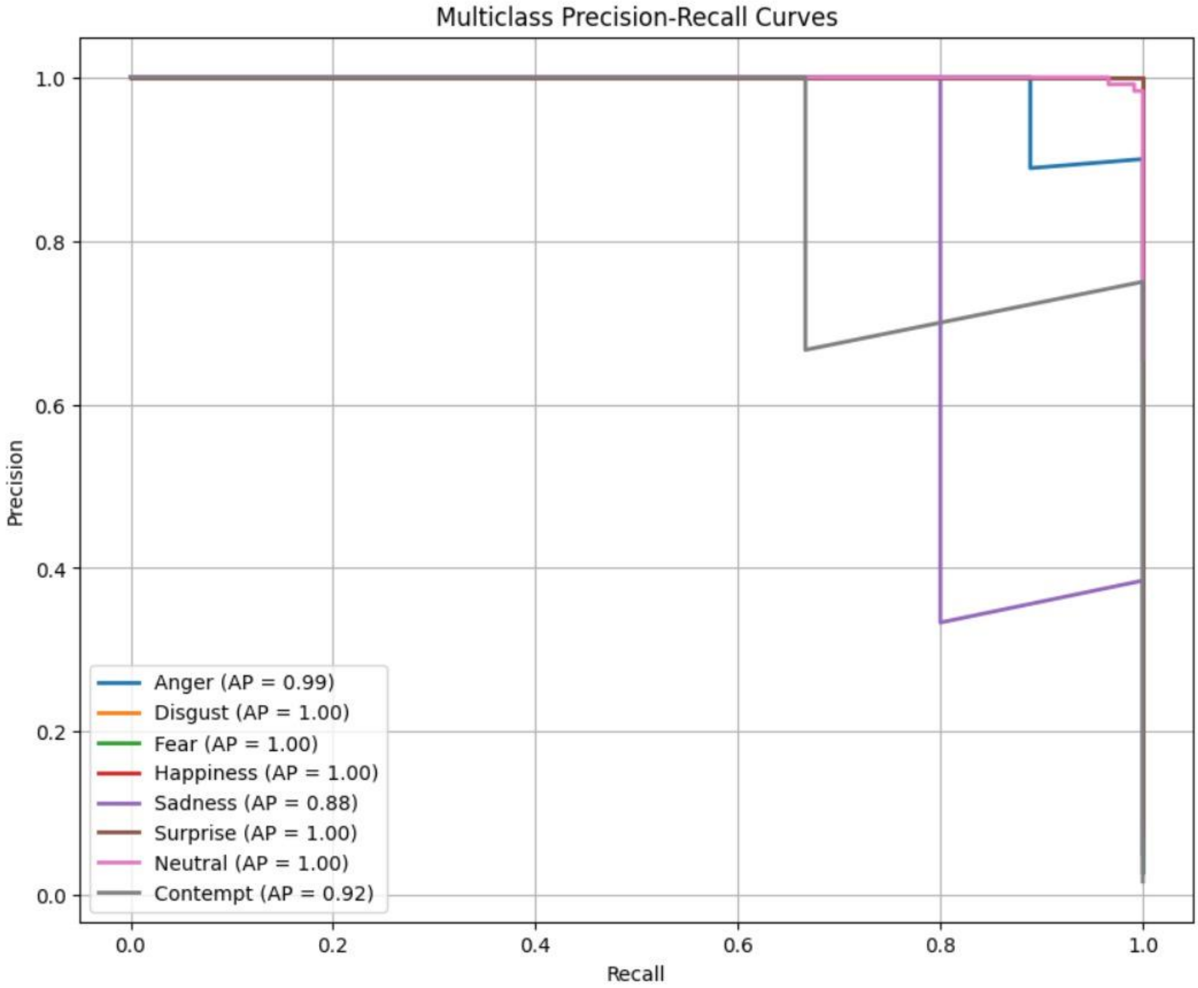


Fig 6. Precision Recall

This plot shows the **multiclass Precision-Recall (PR) curves** for a model predicting eight different emotion classes: **Anger, Disgust, Fear, Happiness, Sadness, Surprise, Neutral,** and **Contempt**. Each curve represents the trade-off between **precision** (y-axis) and **recall** (x-axis) for a specific class, and the **average precision (AP)** score is reported in the legend for each class. We observe that most classes like **Disgust, Fear, Happiness, Surprise,** and **Neutral** achieved an AP of **1.00**, indicating near-perfect precision and recall across all thresholds. **Anger** also performed excellently with an AP of **0.99**. On the other hand, **Sadness** and **Contempt** have relatively lower AP scores of **0.88** and **0.92**, respectively, suggesting that the model finds it slightly more challenging to perfectly distinguish these emotions. The curves for high-AP classes are tightly packed toward the top-right corner, showing strong classifier performance. Meanwhile, the Sadness curve, in particular, dips more significantly, reflecting more variability between precision and recall. Overall, the model demonstrates very high performance, but further improvement could focus on better classifying Sadness and Contempt instances.

CONCLUSION AND FUTURE SCOPE

Conclusion

This research presents a robust facial emotion recognition system utilizing the DenseNet121 architecture applied to the CK+ dataset. By implementing advanced preprocessing techniques, data augmentation, and addressing class imbalance through focal loss and class weighting, our model achieves exceptional performance with 98% accuracy across eight emotion categories. The proposed system offers significant advantages in terms of accuracy, generalization capability, and robustness to class imbalance.

The integration of transfer learning with DenseNet121 provides an efficient approach to feature extraction, while the custom classification layers effectively leverage these features for emotion recognition. The implementation of focal loss and class weighting strategies successfully addresses the challenge of class imbalance, enabling high performance even for under-represented emotion categories.

Future Scope

Several avenues for future research and development can further enhance the capabilities of the proposed system:

1. **Cross-dataset Evaluation**: Testing the model's performance on other emotion recognition datasets such as FER2013, RAF-DB, and KDEF to assess generalization capabilities.
2. **Real-time Implementation**: Optimizing the model for deployment in real-time applications, including mobile devices and embedded systems.
3. **Temporal Emotion Analysis**: Extending the model to analyze emotions in video sequences, capturing the temporal dynamics of facial expressions.
4. **Multi-modal Emotion Recognition**: Integrating facial expression analysis with other modalities such as voice tone, physiological signals, and body language for more comprehensive emotion recognition.
5. **Explainable AI Techniques**: Implementing visualization methods to understand which facial features contribute most significantly to specific emotion classifications.
6. **Attention Mechanisms**: Incorporating attention modules to focus on salient facial regions during classification.
7. **Lightweight Model Variants**: Developing compressed versions of the model suitable for edge computing applications.

The promising results achieved in this study provide a solid foundation for advancing the field of facial emotion recognition, with potential applications spanning healthcare, education, marketing, security, and human-computer interaction domains

REFERENCES

[1] P. Viola and M. J. Jones, “Robust real-time face detection,” *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.

[2] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, “Inception-v4, inception-ResNet and the impact of residual connections on learning,” in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 4278–4284.

[3] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 4700–4708.

[4] P. Ekman and W. V. Friesen, “Facial Action Coding System: A technique for the measurement of facial movement,” Consulting Psychologists Press, Palo Alto, CA, 1978.

[5] P. Lucey et al., “The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, 2010, pp. 94–101.

[6] I. Goodfellow et al., “Challenges in representation learning: A report on three machine learning contests,” in *Neural Information Processing*, Springer, 2013, pp. 117–124. [Includes FER2013 dataset]

[7] S. Jaiswal and A. Prakash, “Facial emotion recognition with convolutional neural networks (FER-CNN),” *Procedia Comput. Sci.*, vol. 132, pp. 89–94, 2018.

[8] T.-Y. Lin et al., “Focal loss for dense object detection,” in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 2980–2988.

[9] D. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in *Proc. Int. Conf. Learn. Representations (ICLR)*, 2015.