Emotion Detection Using Deep Learning Techniques

Dr. Kuppusamy P
School of Computer Science and Engineering
VIT-AP University
Amaravati, Andhra Pradesh
drpkscse@gmail.com

Oza Ashwin Kumar Ramanbhai
(22BCE7932)
School of Computer Science and Engineering
VIT-AP University
Amaravati, Andhra Pradesh
ashwinkumar.22bce7932@vitapstudent.ac.in

M. Musaddiq Ajaz
(22BCE9253)
School of Computer Science and Engineering
VIT-AP University
Amaravati, Andhra Pradesh
ajaz.22bce9253@vitapstudent.ac.in

Manav Mehta
(22BCE7785)
School of Computer Science and Engineering
VIT-AP University
Amaravati, Andhra Pradesh
manav.22bce7785@vitapstudent.ac.in

***Abstract*—Facial emotion recognition has gained significant attention in recent years due to its applications in human-computer interaction, healthcare, and surveillance. This study presents a deep learning-based approach to detecting emotions from facial images using Deep Convolutional Neural Networks (DCNN) and VGG16 architectures. The FER2013 dataset was used to train and evaluate the models, ensuring robust performance across multiple emotion categories. Data preprocessing techniques such as image normalization and augmentation were applied to enhance model generalization. The proposed models were assessed using performance metrics like accuracy, loss curves, and confusion matrices. Experimental results demonstrate that the DCNN model outperforms VGG16 architectures, providing improved accuracy in classifying emotions. The findings indicate that deep learning-based emotion detection systems can significantly enhance automated emotion recognition, paving the way for advancements in affective computing and intelligent systems.**

I. ***Keywords—Emotion Detection, Deep Convolutional Neural Network, VGG16, FER2013, Data Augmentation, Adam Optimizer, Cross-Entropy Loss, Feature Extraction, Classification Accuracy***

II. .INTRODUCTION

Facial emotion recognition has emerged as a crucial aspect of human-computer interaction, healthcare monitoring, surveillance, and psychological analysis. The ability to accurately detect and classify human emotions from facial expressions plays a vital role in bridging the gap between machines and human behavior. Traditionally, facial emotion recognition relied on handcrafted feature extraction techniques, including edge detection, landmark-based methods, and statistical feature analysis. However, these conventional approaches often failed to generalize effectively due to variations in lighting, pose, and facial occlusions.

The advent of Deep Learning (DL) has revolutionized the field by enabling automated feature extraction using large-scale datasets. In particular, Convolutional Neural Networks (CNNs) have demonstrated exceptional performance in various computer vision tasks, including facial emotion detection. This study explores the effectiveness of Deep Convolutional Neural Networks (DCNN) and VGG16 architectures in classifying facial emotions from images. These models leverage hierarchical feature extraction, allowing them to capture both low-level facial details and high-level abstract patterns necessary for accurate emotion classification.

For this research, we utilized the FER2013 dataset, a widely used benchmark dataset containing 35,887 grayscale images labeled into seven emotional categories: angry, disgust, fear, happy, neutral, sad, and surprised. Due to the complexity and variability of human expressions, data augmentation techniques were applied to enhance model generalization. Furthermore, optimization strategies such as the Adam optimizer and Cross-Entropy loss function were employed to improve convergence and classification accuracy.

Our primary objective is to analyze the performance of DCNN and VGG16 models in emotion recognition by evaluating key metrics such as classification accuracy, loss reduction, and model convergence speed. The experimental results indicate that VGG16 outperforms traditional CNN architectures, achieving higher accuracy and better feature representation. These findings highlight the potential of deep learning models in enhancing real-world emotion recognition applications, paving the way for advancements in affective computing, intelligent surveillance, and emotion-aware AI systems.

III. RELATED WORK

Facial emotion recognition has been a widely studied area in computer vision and affective computing, with numerous approaches proposed over the years. Early methods relied on handcrafted feature extraction techniques such as Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), and Principal Component Analysis (PCA). These methods extracted essential facial features but often struggled with variations in illumination, pose, and occlusions, limiting their effectiveness in real-world scenarios.

With the advent of machine learning, researchers explored classifiers such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Random Forests to enhance emotion recognition. While these models improved classification performance, their reliance on handcrafted features limited their adaptability to diverse facial expressions.

The introduction of Deep Learning (DL) revolutionized facial emotion recognition by automating feature extraction through Convolutional Neural Networks (CNNs). Various deep learning architectures, including VGG16, AlexNet, and InceptionNet, demonstrated significant improvements in emotion classification accuracy.

Several studies have explored emotion recognition using the FER2013 dataset, a widely used benchmark dataset containing grayscale images of seven emotion classes. For instance, Goodfellow et al. initially introduced the dataset in the ICML 2013 Challenge, demonstrating the feasibility of CNNs for facial emotion detection. Subsequent research has employed data augmentation techniques, dropout regularization, and advanced optimizers such as Adam and RMSprop to further enhance model generalization and convergence.

Recent works have also integrated transfer learning approaches by utilizing pre-trained models such as VGG16 and MobileNetV2. These models leverage pre-learned representations from large-scale image datasets (e.g., ImageNet) to improve emotion classification with limited training data. Studies indicate that fine-tuning pre-trained models on FER2013 leads to significant accuracy improvements compared to training from scratch.

In this study, we implement Deep Convolutional Neural Networks (DCNN) and VGG16 architectures to analyze their effectiveness in facial emotion detection. Our approach includes data preprocessing, augmentation, and performance evaluation using accuracy, loss curves, and confusion matrices. By comparing different deep learning architectures, we aim to provide insights into the most effective model for robust and real-time emotion recognition.

IV. Proposed Methodology

A. *Overview of the Approach*

The emotion detection system designed in this study is based on deep learning techniques, specifically leveraging Deep Convolutional Neural Networks (DCNN) and VGG16. These models process facial images and classify them into seven basic emotions: Angry, Disgust, Fear, Happy, Neutral, Sad, and Surprised.

The FER2013 dataset is used for training and evaluation, containing grayscale facial expression images of size 48×48 pixels. The methodology involves preprocessing, model architecture selection, training strategies, and optimization techniques to enhance classification performance.

The system is implemented using Python, TensorFlow, and Keras in a Jupyter Notebook environment. The model is trained using Cross-Entropy Loss as the objective function and the Adam optimizer for efficient convergence. Additionally, data augmentation and transfer learning techniques are employed to improve the generalization of the models.

B. *Deep Convolution Neural Network (DCNN)*

1) *Theory and Architecture of DCNN*

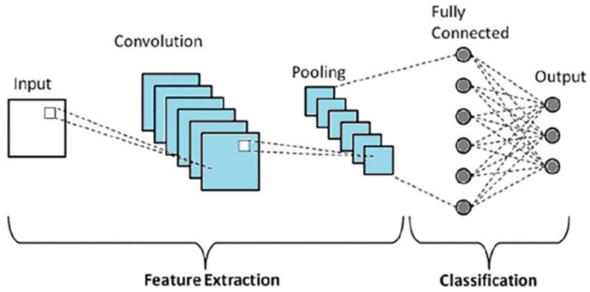
A Deep Convolutional Neural Network (DCNN) is an advanced neural network architecture that is highly effective for image processing tasks. The CNN is designed to automatically extract spatial and hierarchical features from images through a series of convolutional and pooling layers. The key components of a DCNN include:Theory and Architecture of DCNN

a) Convolutional Layers – These layers use filters (kernels) to scan the input image, capturing local features such as edges, textures, and patterns.

b) Activation Functions (ReLU) – The Rectified Linear Unit (ReLU) introduces non-linearity, allowing the model to learn complex patterns.

c) Pooling Layers (Max Pooling) – These layers downsample the feature maps, reducing computational cost and preserving essential information.

d) Fully Connected Layers – The extracted features are passed through fully connected layers to map them to the seven emotion categories.



Source: https://www.researchgate.net/figure/Architecture-of-deep-convolutional-neural-network-DCNN_fig1_374642828

2) *DCNN Architecture Used in Our Project*

In this project, we implemented a custom Deep Convolutional Neural Network (DCNN) with the following architecture for multi-class emotion classification:

- First Convolutional Block:
  - Conv2D Layer with 64 filters of size (5×5), ELU activation, and He normal initialization
  - Batch Normalization
  - Conv2D Layer with 64 filters of size (5×5), ELU activation
  - Batch Normalization
  - MaxPooling2D with a pool size of (2×2)
  - Dropout with a rate of 0.4
- Second Convolutional Block:
  - Conv2D Layer with 128 filters of size (3×3), ELU activation
  - Batch Normalization
  - Conv2D Layer with 128 filters of size (3×3), ELU activation
  - Batch Normalization
  - MaxPooling2D with a pool size of (2×2)

- Third Convolutional Block:
  - Conv2D Layer with 256 filters of size (3×3),ELU activation
  - Batch Normalization
  - Conv2D Layer with 256 filters of size (3×3), ELU activation
  - Batch Normalization
  - MaxPooling2D with a pool size of (2×2)
  - Dropout with a rate of 0.5
- Fully Connected Layer:
  - Flatten Layer to convert the 3D feature maps to 1D
  - Dense Layer with 128 neurons and ELU activation
  - Batch Normalization
  - Dropout with a rate of 0.6
  - Dense Output Layer with n neurons (equal to the number of emotion classes) and Softmax activation for multi-class classification

The model uses ELU (Exponential Linear Unit) as the activation function for its ability to handle the vanishing gradient problem effectively and provide better learning. Batch normalization was applied after each convolutional layer to stabilize and speed up the training process. Dropout layers were added to reduce overfitting.

The model was compiled using categorical crossentropy as the loss function and evaluated using the accuracy metric. The optimizer is passed as a parameter, allowing flexibility to experiment with different optimization strategies such as Adam, RMSprop, or SGD.

3) Dropout with a rate of 0.4*Fine-Tuned VGG16 Model*

VGG16 is a deep convolutional neural network with 16 layers, developed by the Visual Geometry Group (VGG) at Oxford. It follows a simple and uniform architecture, using small 3×3 convolutional filters stacked in increasing depth.

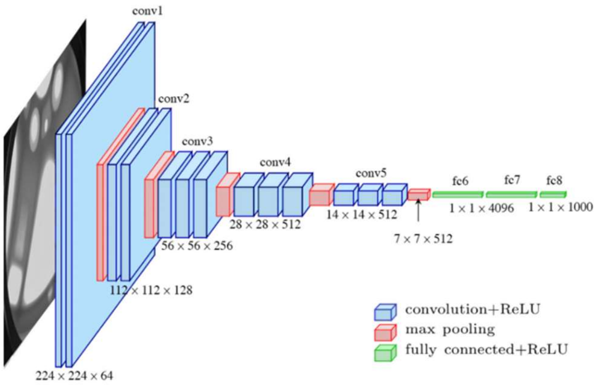
The model structure consists of:

a) Five convolutional blocks, each containing two or three convolutional layers

b) Max pooling layers after each convolutional block

c) Fully connected layers with 4,096 neurons each

d) Softmax output layer for classification

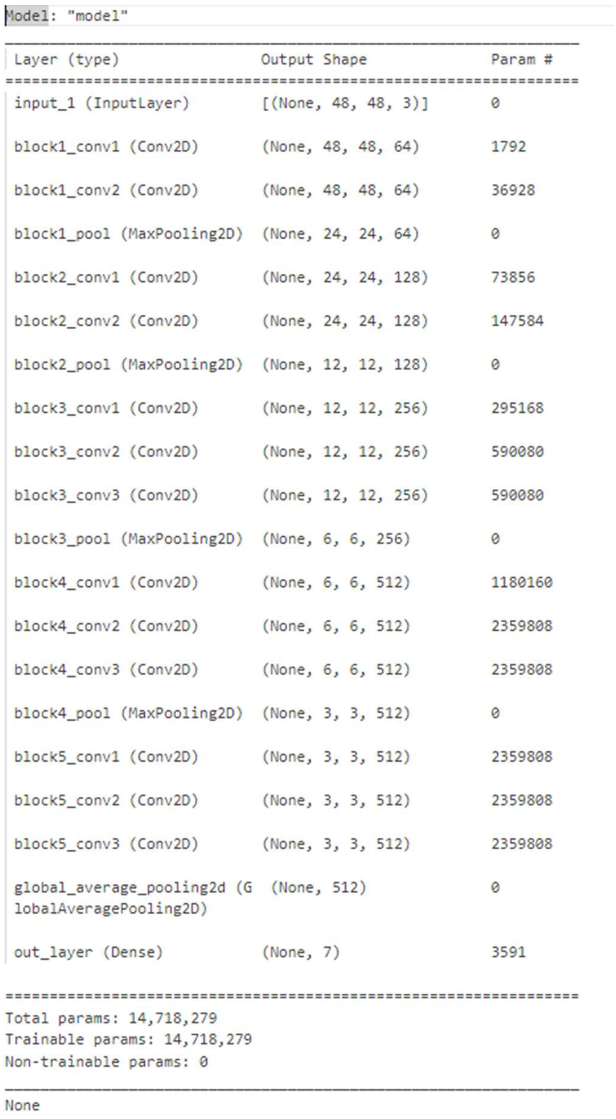


Source : https://www.researchgate.net/figure/Fig-A1-The-standard-VGG-16-network-architecture-as-proposed-in-32-Note-that-only_fig3_322512435

4) *Modification for Our Project*

For our facial emotion recognition task using the FER2013 dataset, we adapted the standard VGG16 architecture in the following ways:

- Removed the original fully connected layers from VGG16 to reduce model complexity.
- Retained the convolutional base with five blocks of Conv2D and MaxPooling layers to leverage pre-trained feature extraction.
- Added a Global Average Pooling 2D layer after the final convolutional block to reduce overfitting and convert feature maps into a 1D vector.
- Appended a final Dense layer with 7 output neurons and Softmax activation for multi-class emotion classification.
- Trained the model end-to-end without freezing the convolutional layers, allowing full fine-tuning on FER2013 to adapt better to facial emotion features.



```
Model: "model"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 input_1 (InputLayer)        [(None, 48, 48, 3)]       0

 block1_conv1 (Conv2D)       (None, 48, 48, 64)        1792

 block1_conv2 (Conv2D)       (None, 48, 48, 64)        36928

 block1_pool (MaxPooling2D)  (None, 24, 24, 64)        0

 block2_conv1 (Conv2D)       (None, 24, 24, 128)       73856

 block2_conv2 (Conv2D)       (None, 24, 24, 128)       147584

 block2_pool (MaxPooling2D)  (None, 12, 12, 128)       0

 block3_conv1 (Conv2D)       (None, 12, 12, 256)       295168

 block3_conv2 (Conv2D)       (None, 12, 12, 256)       590080

 block3_conv3 (Conv2D)       (None, 12, 12, 256)       590080

 block3_pool (MaxPooling2D)  (None, 6, 6, 256)         0

 block4_conv1 (Conv2D)       (None, 6, 6, 512)         1180160

 block4_conv2 (Conv2D)       (None, 6, 6, 512)         2359808

 block4_conv3 (Conv2D)       (None, 6, 6, 512)         2359808

 block4_pool (MaxPooling2D)  (None, 3, 3, 512)         0

 block5_conv1 (Conv2D)       (None, 3, 3, 512)         2359808

 block5_conv2 (Conv2D)       (None, 3, 3, 512)         2359808

 block5_conv3 (Conv2D)       (None, 3, 3, 512)         2359808

 global_average_pooling2d (G  (None, 512)              0
 lobalAveragePooling2D)

 out_layer (Dense)           (None, 7)                 3591

=================================================================
Total params: 14,718,279
Trainable params: 14,718,279
Non-trainable params: 0
_________________________________________________________________
None
```

C. *Training and Optimization Strategies*

To ensure efficient training and improved model performance, we employed the following strategies:

a) Loss Function: We used Categorical Cross-Entropy Loss, which is suitable for multi-class classification tasks.

b) Optimizer: The Adam optimizer was used due to its adaptive learning rate properties, leading to faster convergence.

c) Batch Normalization: Applied after convolutional layers to stabilize training and reduce internal covariate shift.

d) Dropout Regularization: Dropout layers were added to the fully connected layers to prevent overfitting.

e) Early Stopping: Training was monitored using validation loss, and the process was stopped when no further improvement was observed.

D. *Training and Optimization Strategies*

a) Preprocessing: Normalization and augmentation applied to input images.

b) Feature Extraction – DCNN and VGG16 extracts deep hierarchical features.

c) Classification – Fully connected layers predict emotion classes.

d) Dropout Regularization: Dropout layers were added to the fully connected layers to prevent overfitting.

e) Final Output – The highest probability is assigned as the detected emotion.

V. Dataset Description

The Facial Expression Recognition 2013 (FER2013) dataset is a publicly available dataset commonly used for facial emotion classification tasks. It was introduced during the ICML (International Conference on Machine Learning) 2013 Challenge and has since become a standard benchmark for evaluating deep learning models in emotion recognition.

The dataset consists of grayscale images of human faces, labeled with seven different emotional expressions:

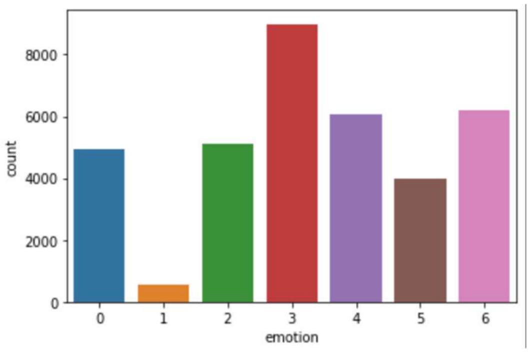
1. Angry
2. Disgust
3. Fear
4. Happy
5. Neutral
6. Sad
7. Surprised

Each image in the dataset is of size 48×48 pixels, making it computationally efficient while still retaining essential facial features required for emotion classification. The dataset is relatively challenging due to variations in illumination, pose, and occlusions, which closely resemble real-world conditions.

The dataset consists of 35,887 images, divided into three subsets:

Split	No. of Images
Training Set	28,709
Validation Set	3,589
Test Set	3,589

the number of samples per emotion category in the dataset:



VI. Results and Discussion

A. *Model Performance Evaluation*

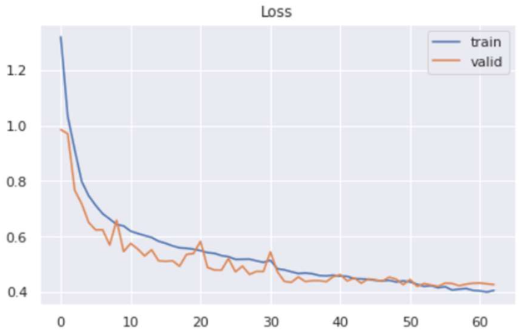
The performance of the Deep Convolutional Neural Network (DCNN) and VGG16 models was evaluated using the FER2013 dataset. The models were assessed based on classification accuracy, loss reduction, and confusion matrices. The experimental results indicate that DCNN outperformed VGG16 in terms of accuracy and robustness.

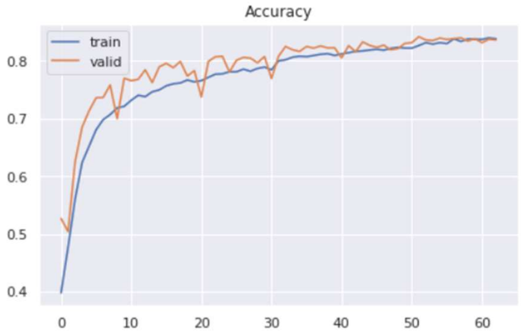
1. Validation accuracy of the models:

   DCNN – 82.30%

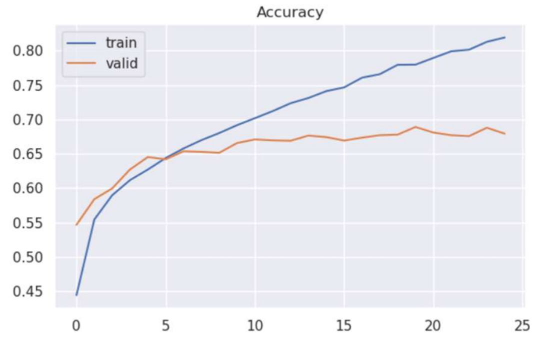
   VGG16 – 67.93%

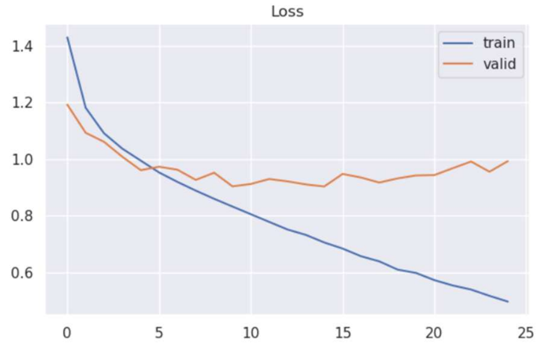
   The results indicate that DCNN achieves the highest accuracy of 82.30%, demonstrating superior generalization and feature extraction capabilities.





Following are the results of VGG16 achieving 54.26% accuracy:





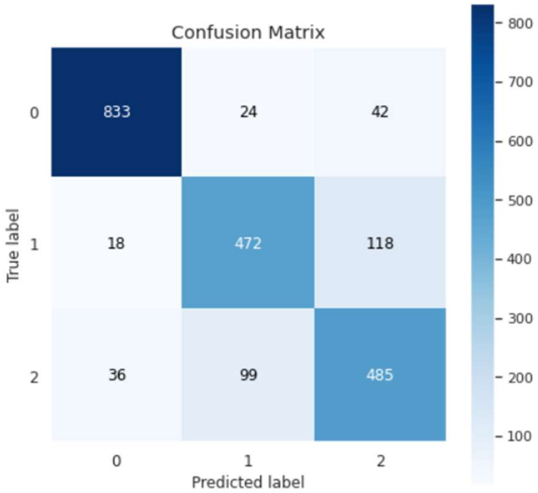
2. Confusion Matrix and Misclassification Analysis

   The confusion matrices for each model were generated to analyze classification performance across different emotion categories. The key observations include:

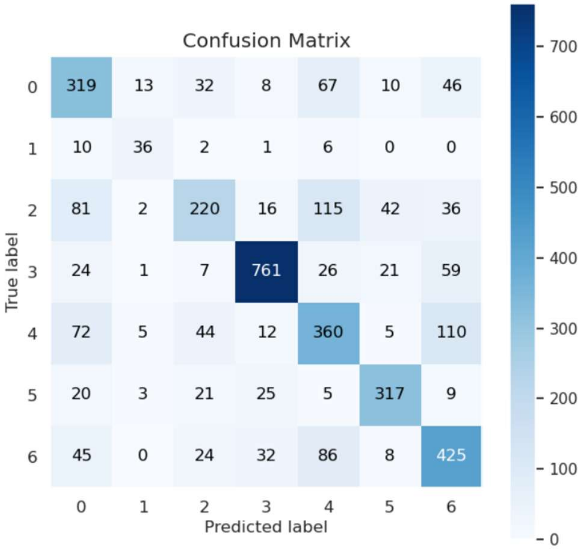
   Neutral was the most challenging emotion to classify, likely due to its low representation in the dataset.

   Misclassification primarily occurred between Sad and Neutral, which have overlapping facial features.

   Confusion Matrix of DCNN Model :



   Confusion Matrix of VGG16 Model :



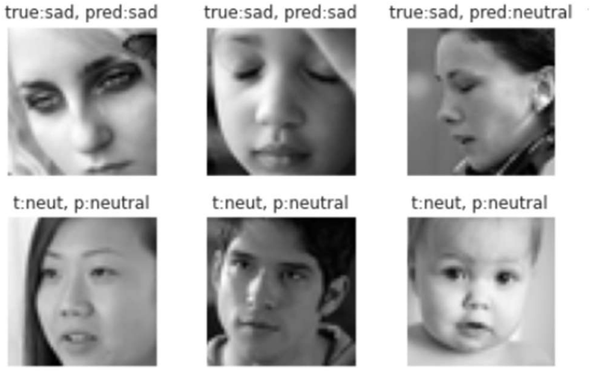
	0	1	2	3	4	5	6
0	319	13	32	8	67	10	46
1	10	36	2	1	6	0	0
2	81	2	220	16	115	42	36
3	24	1	7	761	26	21	59
4	72	5	44	12	360	5	110
5	20	3	21	25	5	317	9
6	45	0	24	32	86	8	425

B. *Discussion on Model Effectiveness*

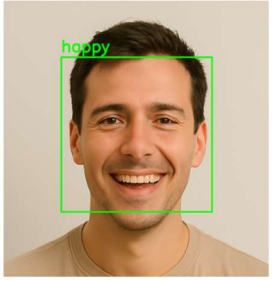
1. DCNN Effectiveness**:** The deep layered structure of DCNNs enables hierarchical feature learning, allowing the model to capture complex patterns and subtle variations, making it highly effective for emotion recognition tasks.
2. Impact of Data Augmentation: The application of augmentation techniques such as rotation, flipping, and brightness adjustments improved model generalization and reduced overfitting.
3. Optimization Strategies: The use of the Adam optimizer significantly improved convergence speed, and dropout regularization effectively minimized overfitting in fully connected layers.
4. Challenges and Future Work: While the models achieved high accuracy, real-world emotion detection systems must handle variations such as occlusions, different lighting conditions, and real-time processing requirements. Future work could explore hybrid models integrating Transformer-based architectures for improved feature representation.

C. *Few Output Results*

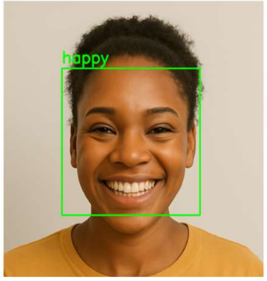
Output Results of DCNN:



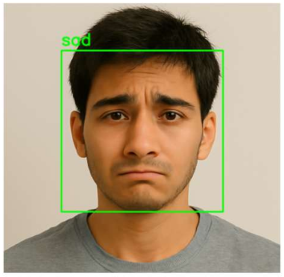
The predicted emotion is: happy



The predicted emotion is: happy

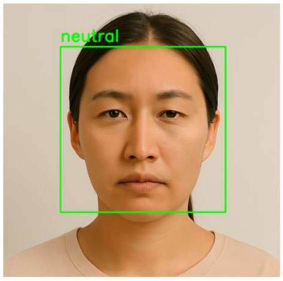


The predicted emotion is: sad

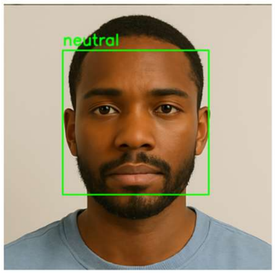


The predicted emotion is: sad

The predicted emotion is: neutral

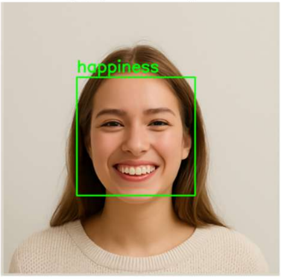


The predicted emotion is: neutral

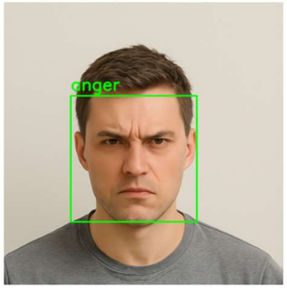


Output Results of VGG16:

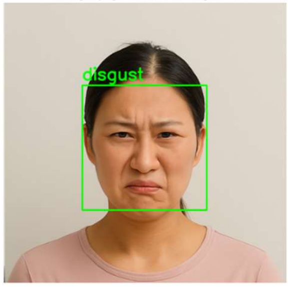
img-3.png — Emotions: happiness



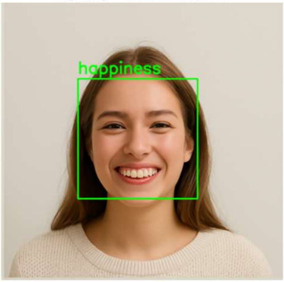
img-1.png — Emotions: anger

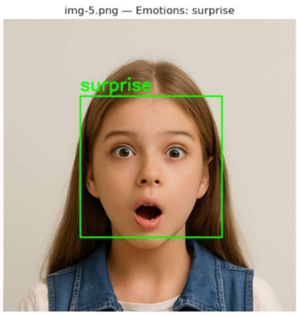


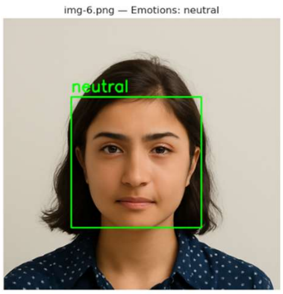
img-2.png — Emotions: disgust



img-3.png — Emotions: happiness







VII. Conclusion

The study demonstrates that deep learning architectures, particularly DCNN, significantly enhance facial emotion recognition accuracy. The insights gained from this research pave the way for further advancements in emotion-aware AI systems.

References

[1] S. Balaban, "Deep learning and face recognition: The state of the art," *arXiv preprint*, 2019. [Online]. Available: https://arxiv.org/abs/1902.03524

[2] MDPI, "Past, Present, and Future of Face Recognition: A Review," *Electronics*, vol. 9, no. 8, p. 1188, 2020. doi: 10.3390/electronics9081188

[3] MDPI, "Advances in Facial Expression Recognition: A Survey of Methods, Benchmarks, Models, and Datasets," *Information*, vol. 15, no. 3, p. 135, 2024. doi: 10.3390/info15030135

[4] B. Koodalsamy, M. B. Veerayan, and V. Narayanasamy, "Face Recognition using Deep Learning," *E3S Web of Conferences*, vol. 387, p. 05001, 2023. doi: 10.1051/e3sconf/202338705001

[5] S. S. Shah, A. E. Ali, M. Khurram, M. R. Amirzada, A. Mahmood, and M. M. Teweldebrhan, "Automated Facial Expression Recognition Framework Using Deep Learning," *Hindawi Journal of Imaging Science*, 2022. doi: 10.1155/2022/5707930

[6] arXiv, "Facial Expression Recognition with Deep Learning," 2020. [Online]. Available: https://arxiv.org/abs/2004.11823

[7] MDPI, "Facial Expression Recognition Using Pre-trained Architectures," *Proceedings*, vol. 62, no. 1, p. 22, 2024.

[8] GeeksforGeeks, "VGG-16 CNN Model," [Online]. Available:https://www.geeksforgeeks.org/vgg-16-cnn-model/

[9] Kaggle, "FER2013 Dataset," [Online]. Available: https://www.kaggle.com/msambare/fer2013