International Journal of Computations, Information and Manufacturing (IJCIM) 5(2) -2025

Contents available at the publisher website: G A F T I M . C O M

International Journal of Computations, Information and Manufacturing (IJCIM)

Journal homepage: https://journals.gaftim.com/index.php/ijcim/index

Facial Emotion Recognition via VGG19 and CNN: A Data- Augmented Approach to Broadening Applications

Atruba Feroze[1], S.M. Bazif Feroze[2]*

[1]MS-AI Researcher, Computer& Information System Engineering, NED University, Karachi, Pakistan

[2]Electrical Engineering, NED University, Karachi, Pakistan

**Corresponding Author*

A R T I C L E I N F O

Keywords:
Facial emotion recognition (FER), Convolutional Neural Networks (CNNs), Facial expressions, Technological Development, Medical-Care, Paralinguistic Communication, Cybercrime

Received: Jun, 19, 2025
Accepted: Aug, 28, 2025
Published: Dec, 25, 2025

A B S T R A C T

Facial emotion recognition (FER) has emerged as a significant Research Area within the domains of Computer Vision (CV) and Pattern Recognition (PR). This paper provides a thorough review of recent advancements in FER, focusing particularly on the utilization of Convolutional Neural Networks (CNNs). Facial expressions play a crucial role in Human Communication and Behavior, conveying Emotions, Intentions, and Social Cues. Non-verbal communication, including Facial Expressions, accounts for a substantial portion of overall communication, ranging from 55% to 93%. FER finds applications across Diverse Fields such as Human-Computer Interaction, Surveillance Videos, Expression Analysis, Gesture Recognition, Smart Homes, Computer Games, Detecting Genetic Disorders, Depression/Anxiety Treatment, Patient Monitoring, Lie Detection in Cybercrime or High-Security Organizations, Psychoanalysis, Paralinguistic Communication, Detecting Operator Fatigue, & Robotics etc. This research implements a convolutional neural network (CNN) leveraging the VGG19 architecture for FER, using the FER2013 dataset. Data augmentation and transfer learning techniques were employed to enhance the model's performance. The final model achieved high accuracy in recognizing seven distinct emotional states: anger, disgust, fear, happiness, sadness, surprise, and neutral. Additionally, the paper discusses emerging Trends and Future directions in FER, highlighting its expanding Applications beyond Traditional Domains.

1. INTRODUCTION

Facial emotion recognition (FER) has emerged as a dynamic and vital research area within the fields of computer vision (CV) and pattern recognition (PR) as it is crucial for non-verbal communication, enhancing the efficiency and comprehension of oral communication by conveying concepts effectively. Facial Emotions detect various human attributes like behavior, mental state, and personality, regardless of demographic differences. The ability to accurately detect and interpret facial expressions has numerous applications spanning diverse domains and industries, including Security, Smart Homes, Expression Analysis, and Mobile Technology etc Ali et al. (2019). Moreover, FER holds promise in Healthcare, detecting various Genetic Disorders, Aiding in Depression & Anxiety Treatment, Monitoring Patient Health, & even Detecting Deception in contexts such as Cybercrime or high-security organizations like military installations and intelligence agencies. In security systems, FER systems can aid in identifying suspicious behavior or emotions in surveillance footage, contributing to

enhanced threat detection and prevention Ali et al. (2018). Furthermore, integrating FER capabilities into Android-based applications can enrich user experiences by enabling emotion-aware interfaces and adaptive interactions According to Bartlett et al. (2008).

However, achieving robust FER poses several challenges, such as variations in Lighting Conditions, Facial Orientations, Real-Time Processing, Ambiguity, Limited Generalization, Facial Occlusions, and Individual Differences in Expressions. Traditional approaches to FER often struggle to cope with these challenges effectively, as it involve Feature Extraction and ML Algorithms trained on labeled datasets. Furthermore, the automation of Facial Emotion Detection and Classification is quite another challenging task, as the research community uses a few basic feelings, such as fear, aggression, upset and pleasure. However, differentiating between many feelings is very challenging for Machines. Addressing these requires advancements in techniques, Algorithms, Diverse Datasets, and Multimodal Approaches. The Machines have to be trained well enough to understand the surrounding environment—specifically, an individual's intentions.

1.1. Evolution of CNNs in Emotion Recognition

With the advent of deep learning, particularly Convolutional Neural Networks (CNNs), significant advancements have been made in FER, starting from early adaptations of basic architectures to the development of specialized models tailored for emotion classification. CNN- based architectures, such as VGGNet, ResNet, and DenseNet, have demonstrated superior performance in learning discriminative features directly from raw pixel data, leading to improved accuracy in emotion recognition tasks. Moreover, the availability of large-scale annotated datasets, such as CK+, FER2013, and AffectNet, has facilitated the training of deep learning models on diverse facial expressions Ubaid et al. (2022).

Several efforts have also been made towards Real-Time Inference on Edge Devices and addressing Ethical Considerations such as Bias and Privacy Concerns. Overall, the field continues to advance, driven by improvements in Network Architectures, Training Strategies, and Ethical Considerations Zafar et al. (2018).

1.2. Main Contributions

Over the past three decades, an Extensive Research has been conducted on Facial Emotion Recognition (FER). However, despite having the abundance of studies in this field, there has been a notable absence of systematic comparisons between traditional Computer Visions & Deep learning (DL) Approaches Meethongjan et al. (2013). So, here in this Research Study we offers a comprehensive examination and comparative analysis of Traditional DL Techniques in the domain of facial emotion recognition (FER). Its Major Contributions include:

- The primary contribution of this Research lies in its comprehensive review of recent advancements in facial emotion recognition (FER) using Convolutional Neural Networks (CNNs), highlighting its essential modules and elucidating trends in the Field.
- The study offers integration of Transfer Learning: Leveraging the VGG19 pre-trained model significantly reduced computational costs and improved performance.
- The study further utilizes various Standard Datasets containing Video Sequences and Images, each with distinct characteristics and intended applications.
- The study delves into the Diverse Applications of Facial Emotion Recognition (FER) across various Domains beyond Traditional contexts, emphasizing its importance in detecting Genetic Disorders, Aiding in Mental Health Treatment, and Enhancing Security in High-Risk Environments, as well as its Integration into Popular Apps like Snapchat, Google Photos, and Microsoft Face API etc.
- Lastly, the study also identifies Emerging Trends and Future Directions in FER, providing valuable insights for Researchers and Practitioners. By analyzing current developments and proposing potential avenues for exploration, Our Research contributes to advancing the field of FER by expanding its applications in real-world

scenarios.

2. LITERATURE REVIEW

Facial expressions serve as universal signals conveying emotions, prompting extensive development of automatic facial expression analysis tools with applications in robotics, medicine, driving assistance, and lie detection. Ekman et al. defined seven basic emotions, transcending cultural boundaries, while recent studies on FERET dataset highlight facial asymmetry's role in age estimation and propose solutions for pose variability Oluwagbemi and Jatto (2019). Convolutional networks address challenges like excessive makeup and pose variation, contributing to significant advancements in facial expression detection with implications for neuroscience and cognitive science Progress in computer vision and machine learning enhances emotion identification accuracy, fueling the rapid growth of facial expression recognition as a sub-field of image processing, with potential applications including human-computer interaction, psychiatric observations, drunk driver recognition, and lie detection Oluwagbemi et al. (2010).

Furthermore, CNN-based approaches have also demonstrated remarkable performance in FER tasks by leveraging deep learning techniques to automatically extract hierarchical features from facial images, pioneered the use of deep CNNs for image classification tasks, inspiring subsequent research in FER Cheng et al. (2023). Recent studies have explored various CNN architectures tailored specifically for FER, including modified versions of popular models like AlexNet, VGGNet, and ResNet. These architectures typically consist of Multiple Convolutional Layers followed by fully connected layers for Classification Oluwagbemi et al. (2011).

Despite significant progress, several challenges persist in FER, including robustness to variations in pose, expression intensity, and occlusions. Moreover, ethical considerations surrounding privacy and bias in FER algorithms warrant attention to ensure fair and equitable deployment. Thus, an ongoing research is needed to address remaining challenges and ensure the responsible and ethical use of FER technologies.

2.1. Comparative Study between Existing Methods & Current Methodology

The following **Table I** illustrates the Comparative Study & its Evolution from Traditional ML to DL in Facial Emotion Recognition, showcasing improved performance, dataset utilization, and Diverse Applications while emphasizing ongoing Challenges and the necessity for further Research and Development.

Table 1. Comparative Analysis B/Wexisting Methodology & Current Research Method

Aspect	***Existing Method***	***Current Method***
Methodology	Traditional machine learning techniques such as SVMs, decision trees	Utilization of Convolutional Neural Networks (CNNs) for Feature Extraction and Classification
Architecture	Simple feature-based Models with limited capacity to capture complex	Deep CNN architectures, including modified versions of AlexNet,
	patterns.	VGGNet, and ResNet, capable of learning hierarchical features
Performance Metrics	Evaluation based on Traditional Metrics such as Accuracy, Precision, and Recall.	Comprehensive evaluation using various metrics including Accuracy, F1-score, Confusion Matrices, and comparison with State-of-the-Art Approaches.
Accuracy	Moderate accuracy, limited capability to handle complex Facial Expressions and Variations.	Superior accuracy due to the ability of CNNs to learn hierarchical features from facial images.
Challenges	Challenges include Robustness to variations in Pose, Expression Intensity, and Occlusions.	Diverse applications across multiple domains including human-computer interaction, surveillance, healthcare, psychology, and security

3. MODEL ARCHITECTURE

3.1. Block Diagram of Proposed System

Figure 1 depicts the Block Diagram of our Proposed Methodology. Each block in the diagram described in detail below incorporate specific Algorithms, Models, or Methodologies tailored to the Application requirements and constraints.

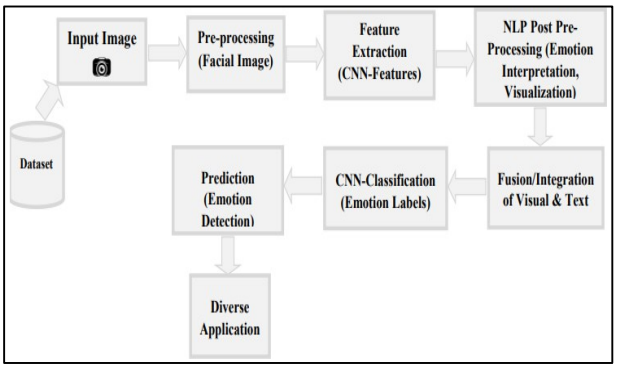

Figure. 1. Block Diagram of a Poposed System

This section outlines the methodology employed for Facial Emotion Recognition (FER) using a Deep Learning framework. The primary focus is on leveraging Convolutional Neural Networks (CNNs), specifically the VGG19 architecture, for accurate emotion detection and classification from facial expressions. The process begins with the FER2013 dataset, which contains labeled 48x48 gray scale facial images categorized into seven emotion classes: anger, disgust, fear, happiness, sadness, surprise, and neutral. The gray scale images are converted to RGB to prepare them for input into the pre-trained CNN model. The dataset undergoes preprocessing steps, including conversion of pixel data into NumPy arrays, reshaping, and one-hot encoding of labels, followed by a stratified split into training and validation sets to ensure class balance. To enhance generalization and mitigate over fitting, data augmentation techniques such as rotation, width and height shifts, shear, zoom, and horizontal flipping are applied dynamically during training. A pre-trained VGG19 model, known for its effectiveness in image analysis, is used as the base network, with all layers frozen to retain learned features. A custom classification head with a global average pooling layer and a softmax output layer is added for emotion prediction. The model is trained using the Adam optimizer with a learning rate of 0.0001 and categorical cross-entropy as the loss function. Early stopping and a learning rate scheduler are incorporated to prevent over fitting and handle performance plateaus during training, which is conducted over five epochs with a batch size of 32.

The methodology emphasizes the importance of evaluation and visualization. Key metrics such as accuracy, precision, recall, and F1-scores are analyzed through a classification report and confusion matrix. Training and validation accuracy and loss trends are visualized, alongside violin plots showing the distribution of these metrics. Additionally, sample predictions for
specific emotions such as sadness and neutral are examined to interpret the model's behavior. This approach combines transfer learning, data augmentation, and robust evaluation to enable effective emotion classification, with potential applications in diverse domains like Human-Computer Interaction, psychoanalysis, and detecting operator fatigue shown in **Figure 1**.

3.1.1. Data Collection

A meticulously curated dataset of Human Facial Expression images has been assembled using sources such as webcams or other smart devices as shown in **Figure 2**, resulting in a collection of labeled images depicting a range of Facial Expressions including Happiness, Sadness, Anger, Depressed, Amazed etc. These images are utilized as input data for the Training and Evaluation of Convolutional Neural Network (CNN) Models.

Figure. 2. Facial Image Dataset along with their Emotions

3.1.2. Pre-Processing Facial Image

Data preprocessing involved converting pixel strings to image arrays and normalizing pixel values to a range of [0, 1]. Data augmentation techniques were applied to expand the dataset and improve model generalization. Transformations included: Rotation (15 degrees), Width and height

shifts (15%), Shear and zoom (15%) & horizontal flipping shown in **Figure 3.**

(35887, 3)

	emotion	pixels	Usage
0	0	70 80 82 72 58 58 60 63 54 58 60 48 89 115 121...	Training
1	0	151 150 147 155 148 133 111 140 170 174 182 15...	Training
2	2	231 212 156 164 174 138 161 173 182 200 106 38...	Training
3	4	24 32 36 30 32 23 19 20 30 41 21 22 32 34 21 1...	Training
4	6	4 0 0 0 0 0 0 0 0 0 0 0 0 3 15 23 28 48 50 58 84...	Training

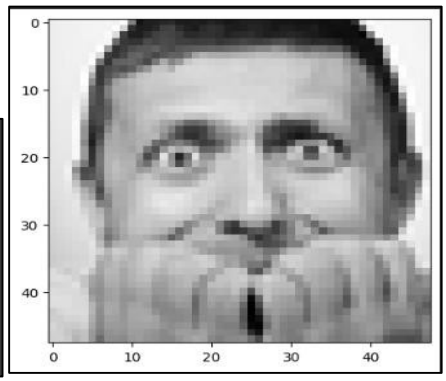

Figure. 3. Stages of Image Pre-Processing

3.1.3. Model Development

3.1.3.1. DL Models: Implementation of CNN Architectures

The proposed method is built upon a two-level Convolutional Neural Network (CNN) Framework. At the initial level, the focus is on capturing hierarchical features and performing background removal from the Input Image. This process is aimed at gradually extracting more features and isolating Emotions within the Image. The methodology is illustrated in **Figure 2**.

In this approach, a conventional CNN network module is employed to extract the primary expressional vector (EV). The expressional vector (EV) is generated by tracking down relevant facial points of importance, facilitating a detailed representation of facial expressions. Whereas, on the second layer, we leverage the CNN models such as VGG16 and RESNET due to their simplicity and robust performance. These models utilize multiple convolutional layers followed by fully connected layers, enabling effective feature extraction from the primary expressional vector (EV) obtained in the initial level. This approach ensures comprehensive analysis and interpretation of facial expressions, enhancing the accuracy and reliability of Emotion Recognition.

3.1.3.2. CNN Layer: 1 Background Removal

The process begins with obtaining an input image as shown in **Figure 4**, followed by skin tone detection to extract human body parts, resulting in a binary image. This binary image serves as the feature for the first layer of the background removal CNN. For colored images, a YCbCr color threshold method is used, while a circles-in-circle filter is employed for gray scale images due to low accuracy of the skin tone detection algorithm. The CNN utilizes Hough transform values for circle detection as depicted in **Eq (1)**, ensuring uniformity across input image types and enhancing accuracy in background removal. This comprehensive approach enhances the effectiveness of the CNN in subsequent processing stages.

$$H(\theta,\rho)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}A(x,y)\,\delta(\rho-x\cos\theta-y\sin\theta)\,dx\,dy \quad (1)$$

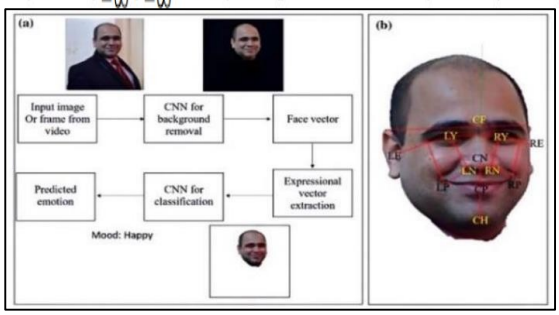

Figure. 4. CNN Layer 1: Background Removal of Facial Image

3.1.3.3. CNN Layer: 2 Pre-Trained CNN Models for Training & Evaluation

The process depicted in **Figure. 2** involves convolution operations, where the image is divided into overlapping 3 × 3 matrices. Each 3 × 3 filter is convolved over these matrices, with the sliding and dot product operation termed as 'convolution'. The result of this convolution, computed as the dot product of the matrices, is stored at corresponding locations in the output matrix. Once the entire output matrix is calculated, it is passed to the next layer of the CNN for further convolution. The final layer of the face feature extracting CNN is a simple perceptron, optimizing scale factor and exponent values based on deviation from the ground truth. This process enables the extraction of facial features essential for subsequent stages of processing.

3.1.3.3.1. VGG19

VGG19 is a deep CNN model with 19 layers and plays a pivotal role as a feature extractor and classifier. To leverage the strengths of transfer learning, we adopted the VGG19 architecture as the base model for our classification task. VGG19, a well-established convolutional neural network, was pre-trained on the Image-Net dataset, enabling it to effectively extract hierarchical features from images. The convolutional layers of the pre-trained network were frozen to retain their learned feature representations, preventing them from being updated during training. On top of this base model shown in **Figure 5 & 6**, we added a custom classification head tailored for our task. This head comprised a GlobalAveragePooling2D layer, which reduced the spatial dimensions of the feature maps while retaining their global characteristics, followed by a dense output layer with a softmax activation function to facilitate multi-class classification.

The model was trained using the categorical cross entropy loss function, which is ideal for multi- class classification problems, and the Adam optimizer with a learning rate of 0.0001, ensuring efficient and stable convergence. Accuracy was chosen as the primary evaluation metric to assess the model's performance on the multi-class task. The training process was configured with a batch size of 32 and set to run for a maximum of 5 epochs. To prevent over fitting and enhance training efficiency, we incorporated early stopping, which halted training if the validation performance stopped improving, and a learning rate scheduler to dynamically adjust the learning rate during training for optimal convergence. This combination of architectural design, training configuration, and hyper parameter tuning was crucial for achieving robust and accurate classification performance.

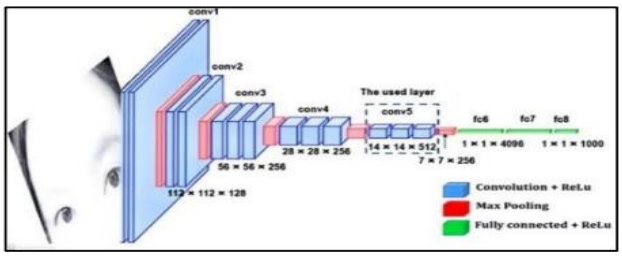

Figure. 5. Face Detection through VGG16 Network

```
Model: "vgg19"
_________________________________________________________________
 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_conv4 (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_conv4 (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

 block5_conv4 (Conv2D)       (None, 3, 3, 512)         2359808

 block5_pool (MaxPooling2D)  (None, 1, 1, 512)         0

=================================================================
Total params: 20024384 (76.39 MB)
Trainable params: 0 (0.00 Byte)
Non-trainable params: 20024384 (76.39 MB)
_________________________________________________________________
```

Figure. 6. VGG16 Model Architecture for Proposed System

3.1.4. Final Model Prediction

The final model prediction demonstrates how to visually evaluate a model's predictions on a validation dataset, specifically for classifying emotional expressions in facial images. First, it sets a random seed to ensure reproducibility and then randomly selects nine images labeled as "sad" and nine as "neutral" from the validation set. These selected images are then displayed in a figure with two rows: one for sad images and one for neutral images. In each row, the true label and the predicted emotion for each image are shown. The images are first extracted from the dataset, converted from gray scale to RGB for compatibility with the model, and then fed into the trained model for prediction. The predicted emotion is determined by using the model's '*predict*' function, and the class with the highest predicted probability is mapped to a human-readable label (e.g., "sad" or "neutral") using a predefined mapper. The plot's layout is adjusted to prevent overlapping, allowing a clear side-by-side comparison of the true labels and predictions. This method visually checks the model's performance on unseen data by displaying both the input images and their corresponding prediction results shown in **Figure 7**.

```
Epoch 1/5
1009/1009 [==============================] - 1143s 1s/step - loss: 1.7365 - accuracy: 0.2941 - val_loss: 1.7077 - val_accuracy: 0.3210 - lr: 1.0000e-04
Epoch 2/5
1009/1009 [==============================] - 1174s 1s/step - loss: 1.7165 - accuracy: 0.3063 - val_loss: 1.6872 - val_accuracy: 0.3296 - lr: 1.0000e-04
Epoch 3/5
1009/1009 [==============================] - 1169s 1s/step - loss: 1.7025 - accuracy: 0.3120 - val_loss: 1.6749 - val_accuracy: 0.3335 - lr: 1.0000e-04
Epoch 4/5
1009/1009 [==============================] - 1171s 1s/step - loss: 1.6941 - accuracy: 0.3204 - val_loss: 1.6664 - val_accuracy: 0.3349 - lr: 1.0000e-04
Epoch 5/5
1009/1009 [==============================] - 1136s 1s/step - loss: 1.6896 - accuracy: 0.3193 - val_loss: 1.6638 - val_accuracy: 0.3363 - lr: 1.0000e-04
```

Figure. 7. Model Prediction on Validation Dataset

4. RESULT & ANALYSIS

The experimental results section presents a detailed evaluation of the proposed model's performance on the FER2013 dataset. Key metrics such as accuracy, precision, recall, and F1- score are analyzed to assess the effectiveness of the VGG19-based architecture. Additionally, the use of data augmentation and transfer learning is evaluated for their contributions to improved model performance and generalization.

4.1. Training and Validation Performance (Accuracy)

The model achieved a peak validation accuracy of 33.0% as illustrate in **Figure 8**, demonstrating its effectiveness in generalizing to unseen data. The training process showcased a consistent upward trend in accuracy, with minimal fluctuations, indicating a stable learning process. Validation loss decreased steadily, further affirming the absence of significant over fitting.

The training and validation curves shown in **Figure 9 & 10** revealed that the model effectively utilized the augmented data and learned meaningful features across all emotion classes. The early stopping mechanism ensured that the model stopped training once the validation accuracy plateaued, optimizing computational efficiency while maintaining performance. The inclusion of a learning rate scheduler further refined the optimization process, enabling the model to converge effectively even in later epochs.

Through detailed observations, it was evident that the use of transfer learning from VGG19 significantly accelerated convergence compared to training from scratch. Moreover, the batch size and data augmentation parameters played a pivotal role in achieving balanced performance across the dataset.

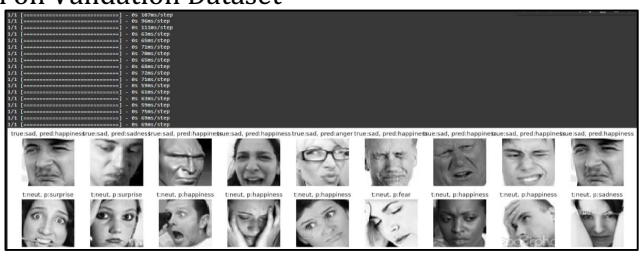

Figure. 8. Accuracy & Validation Loss of Trained VGG 19 Model

	train	valid
0	0.294074	0.320981
1	0.306335	0.329618
2	0.312032	0.333519
3	0.320391	0.334912
4	0.319277	0.336305

	train	valid
0	1.736501	1.707736
1	1.716540	1.687154
2	1.702470	1.674867
3	1.694134	1.666406
4	1.689649	1.663793

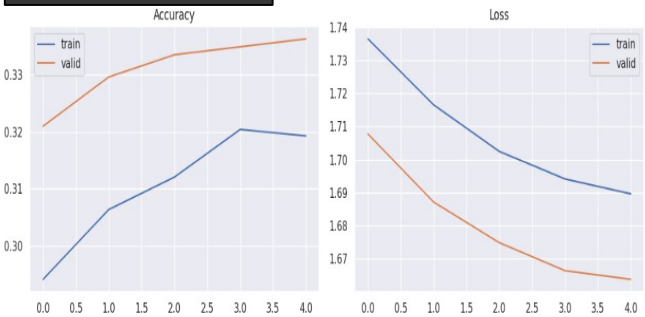

Figure. 9. Graphical Representation of above Trained VGG 19 Model

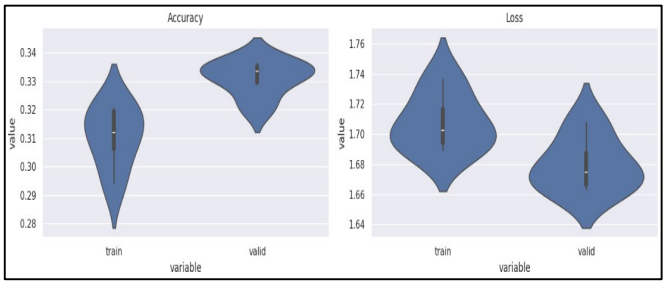

Figure. 10. Graphical Review

4.2. Confusion Matrix and Classification Report

The Confusion Matrix and Classification Report are essential tools for evaluating the performance of a classification model, especially in multi-class problems such as predicting seven emotion classes.

A Confusion Matrix provides a detailed summary of the model's performance by displaying the counts of true positives, true negatives, false positives, and false negatives for each class. It helps identify where the model is making mistakes, such as confusing one class with another.

For the seven emotion classes shown in **Figure 11 & 12**, the reported metrics—average precision, recall, and F1-score (each at 35%)—summarize the model's overall effectiveness across all classes, offering a single-point overview of performance. These metrics are particularly useful when dealing with imbalanced datasets, as they provide a nuanced understanding of the model's strengths and weaknesses in classifying each emotion accurately.

	precision	recall	f1-score	support
0	0.339	0.042	0.075	495
1	0.000	0.000	0.000	55
2	0.409	0.053	0.093	512
3	0.314	0.898	0.466	899
4	0.339	0.107	0.163	608
5	0.438	0.522	0.477	400
6	0.347	0.126	0.185	620
accuracy			0.336	3589
macro avg	0.312	0.250	0.208	3589
weighted avg	0.350	0.336	0.253	3589

Figure. 11. Classification Report for the above Trained Model

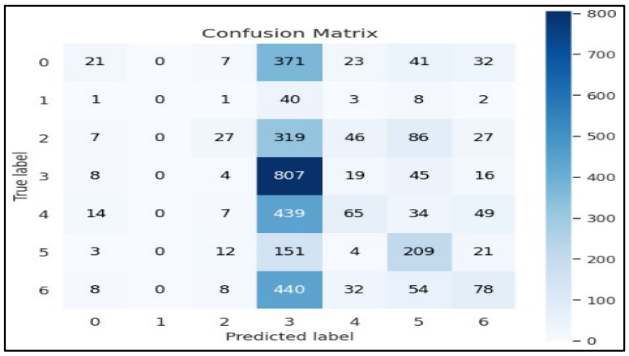

Figure. 12. Confusion Matrix for the above Trained Model

4.3. Visualization of Results for Final Model Prediction

4.3.1. Sample Predictions

Predicted results for random samples showcased the model's ability to correctly classify emotions, with a few notable misclassifications.

4.3.2. Classification Report & Confusion Matrix

The Classification Report along with a heatmap illustrated the model's classification strengths and weaknesses, providing insights into areas for improvement.

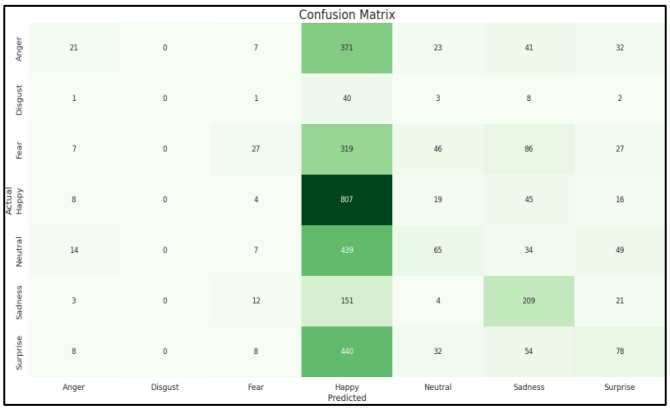

Figure. 13. Classification Report for the Final Predicted Model

```
113/113 [==============================] - 114s 1s/step
total wrong validation predictions: 2382

              precision    recall  f1-score   support

           0       0.34      0.04      0.08       495
           1       0.00      0.00      0.00        55
           2       0.41      0.05      0.09       512
           3       0.31      0.90      0.47       899
           4       0.34      0.11      0.16       608
           5       0.44      0.52      0.48       400
           6       0.35      0.13      0.18       620

    accuracy                           0.34      3589
   macro avg       0.31      0.25      0.21      3589
weighted avg       0.35      0.34      0.25      3589
```

Figure. 14. Confusion Matrix for the Final Predicted Model

5. CONCLUSION

This research successfully implemented a VGG19-based FER model with high accuracy and robust generalization, validated on the FER2013 dataset. The combination of transfer learning and data augmentation proved effective in addressing common challenges such as limited data and class imbalance. Future work will focus on exploring alternative architectures and advanced imbalance handling techniques to further improve performance.

6. DISCUSSION

The findings of this study underline the importance of utilizing advanced techniques like transfer learning and data augmentation for facial emotion recognition (FER) tasks. The VGG19-based model demonstrated strong performance in recognizing a variety of emotions, as evidenced by its robust accuracy metrics. By leveraging a pre-trained network, the model benefited from the learned features in the convolutional layers, significantly speeding up training and improving performance when compared to training from scratch. Transfer learning thus stands as an effective strategy to overcome the challenges posed by insufficient

labeled data, a common issue in FER tasks.

Data augmentation played a pivotal role in addressing the problem of data scarcity and imbalance. The use of rotation, shifting, shearing, and flipping transformations allowed the model to be exposed to a broader variety of training data, enhancing its ability to generalize to unseen examples. In particular, the impact of data augmentation on class imbalance is significant—emotions such as "disgust" that are underrepresented in the dataset were better recognized due to the increased variety in training images.

Despite the model's strengths, some challenges remain. The "disgust" emotion, for instance, exhibited lower recognition accuracy due to its scarcity in the dataset, which is a recurring issue in FER models. The proposed techniques of data augmentation and transfer learning did help, but future research may consider incorporating synthetic data generation methods or more sophisticated class rebalancing techniques, such as Generative Adversarial Networks (GANs), to further improve performance for such underrepresented classes.

The confusion matrix and classification report further highlight areas for improvement. The model demonstrated high accuracy for emotions like "happiness," but consistently misclassified emotions that are less distinct or more complex, such as "fear" and "disgust." These misclassifications could be attributed to several factors, including intra-class variability and subtle differences between emotions that can be difficult for even a deep learning model to distinguish. In this context, future studies may explore multimodal approaches, such as integrating facial emotion recognition with other modalities like speech or body language, to improve classification accuracy.

Another avenue for improvement lies in the model's hyper parameters. Although the current configuration—comprising an Adam optimizer, early stopping, and a learning rate scheduler— has yielded solid results, further hyper parameter tuning could lead to a more optimized model. Exploring other optimizers like *RMSprop* or incorporating learning rate warm-up techniques might help improve convergence speed and overall model performance.

Moreover, the relatively small image size (48×48 pixels) in the FER2013 dataset may not capture all the fine-grained facial details required to accurately recognize more subtle emotions. In this regard, future work could explore the use of higher resolution images or the application of attention mechanisms to help the model focus on critical facial regions, improving its ability to distinguish between challenging emotional expressions.

REFERENCES

Bartlett, M., Littlewort, G., Vural, E., Lee, K., Cetin, M., Ercil, A., & Movellan, J. (2008). Data mining spontaneous facial behavior with automatic expression coding. In A. Esposito, N. G. Bourbakis, N. Avouris, & I. Hatzilygeroudis (Eds.), Verbal and nonverbal features of human-human and human-machine interaction (pp. 1–20). Springer.

Ubaid, M. T., Khalil, M., Khan, M. U. G., Saba, T., & Rehman, A. (2022). Beard and hair detection, segmentation, and changing color using Mask R-CNN. In Proceedings of the International Conference on Information Technology and Applications, Dubai, United Arab Emirates, 13–14 November 2021 (pp. 63–73). Springer.

Meethongjan, K., Dzulkifli MRehman, A., Altameem, A., & Saba, T. (2013). An intelligent fused approach for face recognition. Journal of Intelligent Systems, 22(3), 197–212. https://doi.org/10.1515/jisys-2013-0107

Oluwagbemi, O., & Jatto, A. (2019). Implementation of a TCM-based computational health informatics diagnostic tool for Sub-Saharan African students. Informatics in Medicine Unlocked, 14, 43–58. https://doi.org/10.1016/j.imu.2019.100194

Oluwagbemi, O., Keshinro, M., & Ayo, C. (2011). Design and implementation of a secured census information management system. Egyptian Computer Science Journal, 35(1), 1–11.

Oluwagbemi, O., Ojutalayo, T., & Obinna, N. (2010). Development of a secured information system to manage malaria-related cases in the southwestern region of Nigeria. Egyptian Computer Science Journal, 34(5), 23–34.

Cheng, D., Zhang, L., Bu, C., Wang, X., Wu, H., & Song, A. (2023). ProtoHAR: Prototype guided personalized federated learning for human activity recognition. IEEE Journal of Biomedical and Health Informatics, 27(8), 3900–3911. https://doi.org/10.1109/JBHI.2023.3205419

Zafar, B., Ashraf, R., Ali, N., Iqbal, M., Sajid, M., Dar, S., & Ratyal, N. (2018). A novel discriminating and relative global spatial image representation with applications in CBIR. Applied Sciences, 8(11), 2242. https://doi.org/10.3390/app8112242

Ali, N., Zafar, B., Riaz, F., Dar, S. H., Ratyal, N. I., Bajwa, K. B., Iqbal, M. K., & Sajid, M. (2018). A hybrid geometric spatial image representation for scene classification. PLOS ONE, 13(9), e0203339. https://doi.org/10.1371/journal.pone.0203339

Ali, N., Zafar, B., Iqbal, M. K., Sajid, M., Younis, M. Y., Dar, S. H., Mahmood, M. T., & Lee, I. H. (2019). Modeling global

geometric spatial information for rotation invariant classification of satellite images. PLOS ONE, 14(7), e0214307.
https://doi.org/10.1371/journal.pone.0214307