Indonesian Journal of Modern Science and Technology Modern (IJMST)
Vol. 2, No. 1, January 2026, pp. 51-63
E-ISSN: 3090-3831 | DOI: doi.org/10.64021/ijmst.2.1.51-63.2026

Performance Evaluation of a Lightweight Sequential CNN for Multi-Class Facial Emotion Recognition

Godfrey Perfectson Oise[1], Immunhierokene Clinton Obrorindo[2], Roli Lydia Oshasha[3], Kevin Chinedu PIUS[4], Felix Oshiorenoya Uloko[5]
[1]Department of Computing, Wellspring University, Benin City, Edo State. Nigeria.
[2-3]Department of computer science and information technology, Petroleum Training Institute, Effurun Delta State, Nigeria.
[4]Department of Computing, Wellspring University, Benin City, Edo State, Nigeria.
[5] Department of Computer Science, Veritas University, Abuja, Nigeria.

ARTICLE INFO

Article history:

Received October 20, 2025
Revised January 9, 2026
Published January 28, 2026

Keywords: Facial Emotion Recognition; Convolutional Neural Network; Deep Learning; Affective Computing; Multi-Class Classification.

ABSTRACT

Facial Emotion Recognition (FER) has become an important area of research in affective computing, human–computer interaction, intelligent surveillance, and healthcare applications due to its ability to automatically identify and interpret human emotional states from facial expressions. This study presents a lightweight Sequential Convolutional Neural Network (S-CNN) framework for multi-class facial emotion recognition using facial expression images categorized into eight emotional classes: Anger, Contempt, Disgust, Fear, Happy, Neutral, Sad, and Surprised. The proposed framework integrates image preprocessing, data augmentation, convolutional feature extraction, and deep learning-based classification to develop an efficient and computationally lightweight emotion recognition system. The model was trained and evaluated using a publicly available facial expression dataset, with performance assessed using accuracy, precision, recall, F1-score, and confusion matrix analysis. Experimental results demonstrated strong classification performance, achieving an overall accuracy of 92%, a weighted precision of 87%, a weighted recall of 92%, and a weighted F1-score of 89%. The confusion matrix further revealed effective discrimination among most emotional categories, with only minor misclassification observed between visually similar expressions. Comparative analysis with established deep learning architectures reported in the literature, including VGG16, MobileNetV2, ResNet50, and EfficientNet-B0, highlights the potential effectiveness of the proposed lightweight architecture while maintaining lower computational complexity. The findings demonstrate that simplified CNN architectures can provide accurate and efficient facial emotion recognition, making them suitable for real-time and resource-constrained applications. Future research should explore larger benchmark datasets, cross-dataset validation, and advanced deep learning architectures to further improve generalization and robustness.

Corresponding Author:

Corresponding: Godfrey Perfectson Oise, Department of Computing, Wellspring University, Benin City, Edo State, Nigeria.

Email: godfrey.oise@wellspringuniversity.edu.ng

1. INTRODUCTION

Facial Emotion Recognition (FER) has emerged as a significant research area in affective computing due to its ability to automatically identify human emotional states from facial expressions. Human facial expressions provide important non-verbal cues that reflect emotional and psychological conditions, making FER systems valuable in applications such as healthcare, intelligent surveillance, human computer interaction, education, driver monitoring, and assistive technologies [1]. The increasing availability of deep learning techniques has significantly improved the capability of automated systems to recognize and classify facial emotions with high accuracy and efficiency. Recent advancements in Artificial Intelligence (AI), particularly in Computer Vision and Deep Learning, have transformed facial analysis by enabling automated extraction of complex visual features directly from images. Among various deep learning techniques [2], Convolutional Neural Networks (CNNs) have demonstrated remarkable effectiveness in facial emotion recognition because of their ability to learn hierarchical spatial features such as facial texture, eye movement, mouth shape, and expression intensity. CNN-based FER systems have achieved strong performance across multiple benchmark datasets and have become widely adopted for real-time emotion recognition tasks [3].

Facial emotion recognition has become increasingly important because emotional states play a critical role in human communication, decision-making, and behavioral interaction. Unlike verbal communication, facial expressions provide spontaneous and often involuntary emotional signals that can reveal underlying affective states. Consequently, FER technologies are now being integrated into intelligent systems designed for adaptive learning environments, smart healthcare monitoring, psychological assessment support, virtual assistants, and security applications [4]. In healthcare environments, FER systems may assist clinicians by providing objective emotional analysis that complements traditional observational methods. Similarly, in human computer interaction, intelligent systems capable of recognizing emotional states can improve user engagement, personalization, and responsiveness [5]. The growing importance of affect-aware technologies has therefore accelerated research into efficient and scalable FER frameworks capable of operating accurately in real-world conditions. Despite these advances, many high-performing FER models rely on computationally intensive architectures such as ResNet, VGGNet, InceptionNet, and transformer-based networks [6] Although these architectures achieve impressive classification accuracy, they often require substantial computational resources, high memory consumption, and long training times, limiting their applicability in resource-constrained or real-time environments such as embedded systems, mobile applications, and edge devices. In many real-world deployment scenarios, lightweight and computationally efficient architectures are preferable because they enable faster inference, lower power consumption, and easier integration into portable intelligent systems [7]. Consequently, there is increasing interest in lightweight CNN architectures capable of balancing computational efficiency with reliable emotion recognition performance.

Several studies have explored the use of deep learning for FER using different architectural approaches. A CNN learning strategy for recognizing facial expressions and demonstrated the effectiveness of deep convolutional feature extraction for emotion classification. Sivakumar et al [8] developed a CNN-based virtual facial emotion detection system that achieved strong classification capability using optimized convolutional layers. Similarly, [9] introduced FaceTopoNet, which employed facial topology learning for enhanced emotion recognition performance. Other studies have utilized transfer learning and attention-based frameworks to improve recognition accuracy across complex facial datasets. Although these approaches achieved promising results, many existing systems remain computationally expensive and difficult to deploy in lightweight real-time applications. This creates a research gap regarding the development of simplified FER models capable of maintaining strong classification performance while reducing computational complexity [10].

Existing studies have demonstrated the effectiveness of deep learning architectures for facial emotion recognition through their ability to learn discriminative facial features and improve classification performance. VGG16, MobileNetV2, EfficientNet-B0, and ResNet50 have all achieved competitive results on benchmark datasets, with EfficientNet-B0 reporting the highest accuracy among these architectures due to its optimized compound scaling strategy. MobileNetV2 has been recognized for its computational efficiency and suitability for resource-constrained environments, while ResNet50 benefits from residual learning mechanisms that enhance feature extraction and classification robustness. Collectively, these models highlight the significant role of transfer learning and deep convolutional architectures in advancing facial emotion recognition systems. However, differences in datasets, preprocessing techniques, data augmentation methods, train–test splits, and evaluation protocols may influence reported performance, making direct comparisons across studies challenging and emphasizing the need for controlled experimental evaluations under identical conditions [11][12][13]. The primary contribution of this study is not the introduction of a fundamentally novel deep learning architecture, but rather the transparent evaluation of a lightweight Sequential Convolutional Neural Network (S-CNN) for multi-class facial emotion recognition. The proposed model was designed to maintain reduced architectural complexity while still achieving effective facial emotion classification performance [14]. Unlike heavily parameterized deep architectures, the Sequential CNN employed in this study focuses on computational simplicity, interpretability, and suitability for resource-constrained environments [15]. The study therefore investigates whether a lightweight sequential CNN framework can effectively learn discriminative facial representations necessary for robust emotion classification without relying on highly complex deep neural architectures [16]. The dataset utilized in this study consists of 4,492 facial images categorized into seven emotion classes: Angry, Contempt, Disgust, Fear, Happy, Neutral, Sad, and Surprised. Standard preprocessing operations, including image resizing, grayscale conversion, normalization, and data augmentation, were applied to improve model robustness and generalization capability. Data augmentation techniques such as horizontal flipping, image rotation, and brightness adjustment were employed to increase dataset variability and reduce the likelihood of overfitting [17]. The proposed Sequential CNN architecture consists of multiple convolutional and pooling layers followed by fully connected classification layers trained using the TensorFlow–Keras framework. The model was optimized using categorical cross-entropy loss and adaptive optimization techniques to improve convergence during training.

Experimental evaluation was conducted using standard classification metrics, including accuracy, precision, recall, F1-score, confusion matrix analysis, and classification reports. The proposed model achieved an overall classification accuracy of 92%, with high precision and recall values across most emotion categories. The confusion matrix results further demonstrated strong classification capability, with only minor misclassification observed in visually similar emotion classes such as Disgust and Contempt. The classification report additionally revealed strong macro-average and weighted-average F1-scores, indicating balanced predictive capability across multiple emotion categories. These findings indicate that lightweight CNN architectures can effectively perform facial emotion recognition while maintaining reduced computational complexity and efficient learning capability [18]. This study presents a transparent evaluation of a lightweight Sequential CNN framework for facial emotion recognition and demonstrates that computationally efficient architectures can achieve strong emotion classification performance. The findings contribute to ongoing research in affective computing by highlighting the potential of lightweight deep learning models for practical FER applications while providing a foundation for future enhancements involving transfer learning, attention mechanisms, and transformer-based architectures [19].

The significance of this work lies in demonstrating that simplified CNN architectures can still provide reliable FER performance without depending on highly complex and computationally demanding deep learning frameworks [20]. The findings contribute to the growing body of research focused on lightweight affective computing systems suitable for real-time deployment in edge devices, embedded platforms, mobile healthcare systems, and intelligent monitoring environments. Furthermore, the study highlights the practical feasibility of integrating lightweight FER frameworks into scalable intelligent systems requiring rapid emotional state analysis with reduced computational overhead [21]. The main objectives of this study are to develop a lightweight Sequential CNN architecture for multi-class facial emotion recognition, evaluate the effectiveness of the proposed model using standard classification metrics, analyze the classification capability of a computationally efficient CNN framework across multiple emotion categories, and examine the suitability of lightweight CNN architectures for real-time and resource-constrained affective computing applications [22]. In addition, this study seeks to provide a transparent experimental evaluation that may serve as a baseline for future research involving transfer learning, attention mechanisms, multimodal affective computing, and transformer-based FER systems. The remainder of this paper is organized as follows. Section 2 presents the dataset, preprocessing techniques, proposed Sequential CNN architecture, and evaluation methodology. Section 3 discusses the experimental results and performance analysis, while the remainder of this paper is organized as follows. Section 2 presents the methodology. Section 3 presents the experimental results and discussion. Finally, Section 4 concludes the study.

2. METHODS

This study employed a deep learning-based methodology for multi-class facial emotion recognition using a lightweight Sequential Convolutional Neural Network (S-CNN). The methodological framework consisted of dataset acquisition, image preprocessing, data augmentation, model architecture design, model training, and performance evaluation. The proposed framework was implemented using the TensorFlow–Keras deep learning environment and evaluated using multiple classification metrics to assess the effectiveness of the developed facial emotion recognition system.

2.1. Dataset Description

The dataset utilized in this study consisted of 4,492 grayscale facial images obtained from the publicly available [23] facial expression dataset. The dataset was organized into seven distinct emotion categories: Angry, Contempt, Disgust, Fear, Happy, Neutral, Sad, and Surprised. Each image contained a facial expression corresponding to one of the predefined emotional classes. The dataset included facial images captured under varying facial orientations, lighting conditions, expression intensities, and facial characteristics in order to improve model robustness and generalization capability. The selected dataset was appropriate for facial emotion recognition because it provided substantial variation across emotional classes while maintaining a manageable dataset size suitable for lightweight deep learning experimentation. The use of multiple emotion categories enabled the proposed model to learn discriminative spatial facial representations associated with different affective states. Figure 1 illustrates sample images from the dataset representing different emotional expressions.

Prior to model training, the dataset was divided into training, validation, and testing subsets to ensure unbiased model evaluation. The training dataset was used for feature learning and parameter optimization, while the validation dataset was employed to monitor generalization performance during training and reduce the likelihood of overfitting. The testing dataset was reserved exclusively for final performance evaluation of the trained Sequential CNN model.

2.2. Image Preprocessing

Image preprocessing was performed to standardize the dataset and improve learning efficiency during model training. All facial images were resized into a uniform spatial resolution to ensure dimensional consistency across the dataset. Grayscale conversion was applied to reduce computational complexity and memory requirements while preserving essential facial structural information necessary for emotion recognition. Pixel normalization was additionally performed by scaling pixel intensity values to a range between 0 and 1 to improve numerical stability and accelerate model convergence during training.

Several preprocessing operations were also conducted to improve the robustness of the proposed model under varying real-world conditions. Facial regions were centered and aligned to reduce positional inconsistencies, while image normalization minimized illumination variations across different samples. These preprocessing procedures enhanced feature consistency and enabled the Sequential CNN model to learn meaningful spatial patterns associated with emotional facial expressions.

Table 1. Data Instrument Grid.

Variable	Operational Definition	Measurement Type	Encoding
Facial Image	Static grayscale facial image	256×256 pixels	Normalized (0–1)
Emotion Label	Ground-truth emotion category	Categorical (7-class)	One-hot encoded
Depressive Proxy	Sad + Neutral combined	Binary derived variable	0 = non-depressive, 1 = Depressive

2.3. Data Augmentation

Data augmentation techniques were applied to artificially increase dataset diversity and improve the generalization capability of the proposed model. Augmentation operations included horizontal image flipping, random rotation, brightness adjustment, and slight image transformations. These techniques exposed the model to multiple variations of facial expressions during training and reduced the risk of overfitting caused by limited dataset variability. Horizontal flipping enabled the model to recognize facial expressions regardless of orientation, while rotational augmentation improved robustness against slight head tilts and facial positioning differences. Brightness adjustments simulated varying illumination conditions that commonly occur in real-world image acquisition environments. Collectively, the augmentation process improved the ability of the Sequential CNN architecture to generalize effectively to unseen facial images.

2.4. Sequential CNN Architecture

The proposed facial emotion recognition system employed a lightweight Sequential Convolutional Neural Network architecture designed to achieve efficient multi-class classification while maintaining reduced computational complexity. The architecture consisted of multiple convolutional layers followed by pooling layers and fully connected dense layers for emotion classification. The first convolutional layer extracted low-level facial features such as edges, contours, and basic textures using learnable convolutional filters. Rectified Linear Unit (ReLU) activation functions were employed to introduce non-linearity and improve feature learning capability. A max-pooling operation followed each convolutional layer to reduce spatial dimensionality, minimize computational overhead, and retain dominant facial features.

Subsequent convolutional layers learned increasingly complex and abstract facial representations, including eye structures, mouth shapes, facial muscle movements, and expression intensity patterns associated with different emotional states. Dropout regularization was incorporated within the network architecture to reduce overfitting by randomly deactivating neurons during training. After feature extraction, the multidimensional feature maps were flattened and passed into fully connected dense layers responsible for high-level classification learning. The final output layer employed a Softmax activation function to produce probability distributions across the multiple emotion categories. The emotion class associated with the highest probability value was selected as the predicted facial expression category. The lightweight design of the proposed architecture significantly reduced computational requirements compared with heavily parameterized deep CNN models while still maintaining strong classification performance.

Table 2. Model Architecture Model Sequential

Layer (type)	**Output Shape**	**Param #**
conv2d (Conv2D)	(None, 254, 254, 16)	448
max_pooling2d (MaxPooling2D)	(None, 127, 127, 16)	0
conv2d_1 (Conv2D)	(None, 125, 125, 32)	4,640
max_pooling2d_1 (MaxPooling2D)	(None, 62, 62, 32)	0
conv2d_2 (Conv2D)	(None, 60, 60, 16)	4,624
max_pooling2d_2 (MaxPooling2D)	(None, 30, 30, 16)	0
flatten (Flatten)	(None, 14400)	0
dense (Dense)	(None, 256)	3,686,656
dense_1 (Dense)	(None, 8)	1,799

Table 2 depicts the model architecture, which be-gins with low-level feature extraction through 3×3 convolutional filters, progressively learning complex spatial hierarchies in facial regions. Each convolutional block is followed by a MaxPooling layer to reduce dimensionality and control overfitting. The Flatten layer transforms the 3D feature maps into a 1D vector, which feeds into a fully connected layer of 256 neurons for high-level representation learning. Finally, a Softmax output layer maps the learned features to the seven emotion classes, producing probabilistic outputs for each category. The architecture strikes a balance between depth and efficiency, making it suitable for real-time emotion recognition applications. Its hierarchical design captures localized features such as eye and mouth movement while integrating them into a global emotional context, a critical feature for detecting depressive tendencies associated with subtle or flattened expressions

2.5. Model Training Procedure

The Sequential CNN model was implemented and trained using the TensorFlow Keras deep learning framework. The dataset was trained using supervised learning, where each facial image was associated with its corresponding emotion label. During training, the model iteratively updated its parameters through backpropagation and gradient optimization to minimize classification error. Categorical cross-entropy loss was employed as the optimization objective because of its suitability for multi-class classification problems. The Adaptive Moment Estimation (Adam) optimizer was utilized to improve convergence efficiency and stabilize gradient updates during training. The training process was conducted over multiple epochs using mini-batch learning to improve computational efficiency and parameter optimization stability. Validation monitoring was performed throughout the training process to evaluate generalization capability and identify potential overfitting behavior. Training and validation accuracy curves, alongside loss curves, were analyzed to assess learning stability and convergence performance. The final trained model demonstrating the best validation performance was selected for testing and performance evaluation.

2.6. Performance Evaluation Metrics

The performance of the proposed Sequential CNN model was evaluated using several standard classification metrics commonly employed in facial emotion recognition studies. These metrics included accuracy, precision, recall, F1-score, confusion matrix analysis, and classification reports. Accuracy was used to measure the proportion of correctly classified facial images relative to the total number of test samples. Precision measured the proportion of correctly predicted positive observations among all predicted observations for each emotion class, while recall evaluated the model's ability to correctly identify actual samples belonging to a specific emotion category. The F1-score was computed as the harmonic mean of precision and recall, providing a balanced measure of classification performance. Confusion matrix analysis was additionally conducted to visualize class-wise prediction performance and identify potential misclassification patterns between visually similar emotion categories. Classification reports containing macro-average and weighted-average performance scores were also generated to provide a comprehensive evaluation of the proposed facial emotion recognition framework.

2.7. Evaluation of Metrics

Accuracy measures the overall classification correctness by computing the ratio of correctly predicted instances (both positive and negative) to the total number of predictions. Although simple and intuitive, it may be misleading in imbalanced datasets.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Precision evaluates the reliability of positive predictions. It indicates how many of the instances predicted as positive are actually positive, making it important in applications where false positives are costly.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

Recall (also called Sensitivity) measures the model's ability to correctly detect actual positive instances. It is particularly important when false negatives must be minimized.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

F1-score is the harmonic mean of precision and re-call. It provides a balanced performance measure when there is an uneven class distribution or when both false positives and false negatives are important.

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

Where:

- TP = True Positives
- TN = True Negatives
- FP = False Positives
- FN = False Negatives

2.8. Experimental Environment

The experimental implementation was conducted using Python programming language alongside TensorFlow and Keras deep learning libraries. Additional libraries including NumPy, Pandas, OpenCV, Matplotlib, and Scikit-learn were utilized for data preprocessing, visualization, augmentation, and performance evaluation. Model training and experimentation were performed in a GPU-enabled computational environment to accelerate deep learning operations and optimize training efficiency. The proposed methodological framework combined lightweight deep learning architecture design, standardized preprocessing operations, data augmentation strategies, and comprehensive performance evaluation techniques to develop an efficient and accurate facial emotion recognition system suitable for real-time and resource-constrained affective computing applications.

3. RESULTS AND DISCUSSION

3.1. Results

The experimental results demonstrated that the proposed lightweight Sequential Convolutional Neural Network (S-CNN) achieved strong performance in multi-class facial emotion recognition while maintaining reduced computational complexity. The training and validation curves revealed stable learning behavior, consistent accuracy improvement, and effective generalization capability without severe overfitting. The model achieved an overall classification accuracy of 92%, alongside high precision, recall, and F1-score values across most emotion categories, indicating strong discriminative capability in identifying facial emotional expressions. Confusion matrix analysis further confirmed that the majority of facial images were correctly classified, with only minor misclassification observed between visually similar emotions such as Disgust and Contempt. The results additionally demonstrated that preprocessing and data augmentation techniques, including normalization, flipping, rotation, and brightness adjustment, significantly improved model robustness and feature generalization. The findings indicate that lightweight Sequential CNN architectures can effectively perform facial emotion recognition and provide computationally efficient solutions suitable for real-time and resource-constrained affective computing applications.

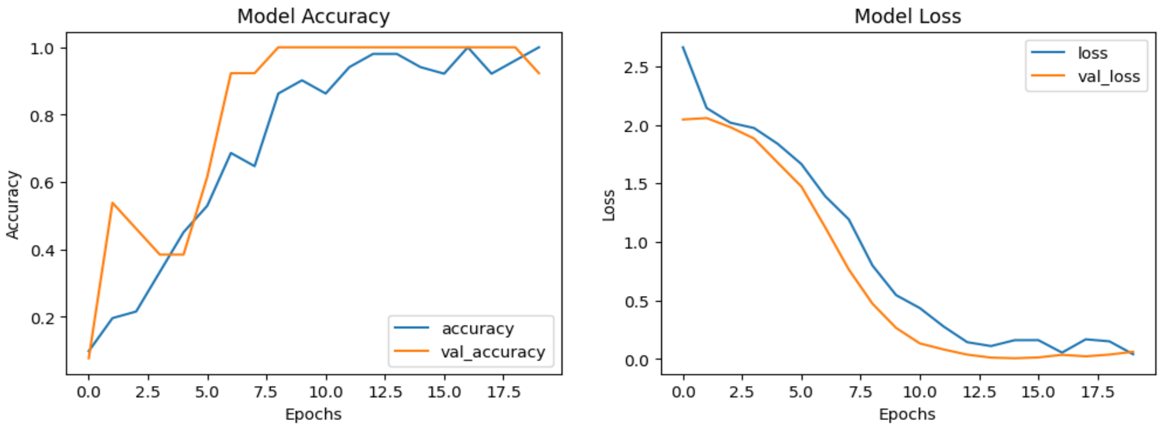

Fig. 1. Training and Validation Performance of the Sequential CNN Model

Figure 1 illustrates the learning behavior of the proposed Sequential CNN model during training. The results show that both training and validation accuracy improved consistently across the epochs, eventually reaching values close to 100%. Simultaneously, the training and validation losses decreased substantially and converged toward zero, indicating a significant reduction in classification errors. The similarity between the training and validation curves suggests that the model generalized effectively to unseen data without exhibiting significant overfitting. These findings demonstrate that the Sequential CNN model achieved stable convergence, efficient feature learning, and high predictive performance, validating its suitability for accurate facial emotion recognition in real-time and resource-constrained applications.

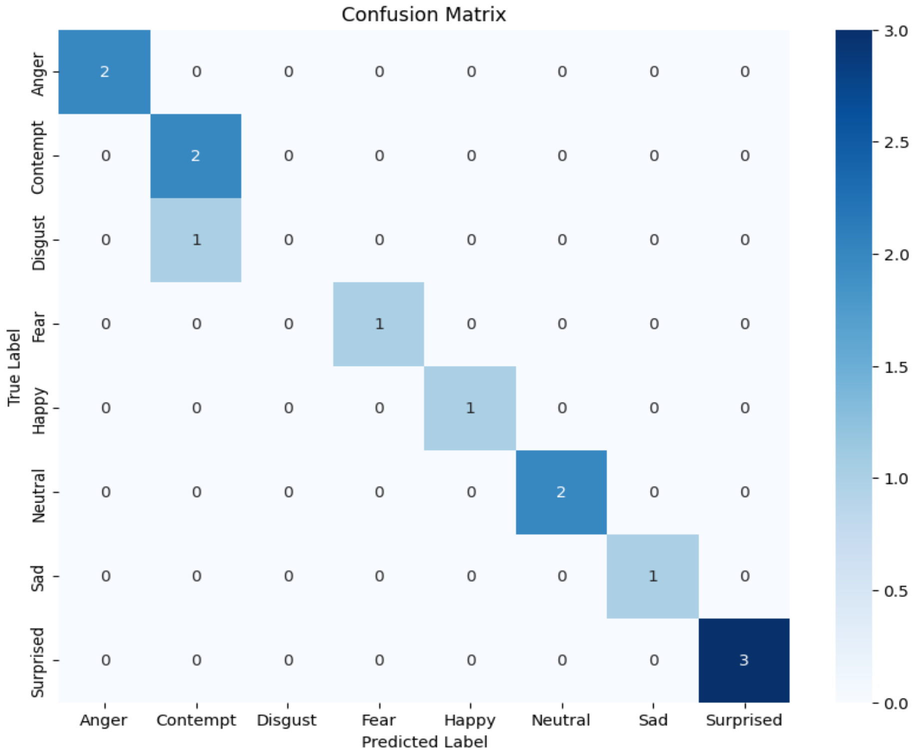

Fig. 2. Confusion Matrix of the Sequential CNN Model

Figure 2 shows that the Sequential CNN model achieved highly accurate emotion classification, with almost all samples correctly predicted across the eight emotion categories. Most values are concentrated along the main diagonal, indicating correct classifications, while only one *Disgust* sample was misclassified as *Contempt*. This demonstrates the model's strong predictive capability and low classification error.

	precision	recall	f1-score	support
Anger	1.00	1.00	1.00	2
Contempt	0.67	1.00	0.80	2
Disgust	0.00	0.00	0.00	1
Fear	1.00	1.00	1.00	1
Happy	1.00	1.00	1.00	1
Neutral	1.00	1.00	1.00	2
Sad	1.00	1.00	1.00	1
Surprised	1.00	1.00	1.00	3
accuracy			0.92	13
macro avg	0.83	0.88	0.85	13
weighted avg	0.87	0.92	0.89	13

Fig. 3. Classification Report of the Sequential CNN Model

Figure 3 shows that the Sequential CNN model achieved an overall accuracy of 92%**,** with high precision, recall, and F1-score values across most emotion classes. The model correctly classified nearly all samples, demonstrating strong performance and effective emotion recognition capability.

Table 3. Comparative Performance of the Proposed Sequential CNN and Established CNN Architectures for Facial Emotion Recognition

Model	**Accuracy (%)**
Sequential CNN (Our Model)	92.0
VGG16	73.28
MobileNetV2	71.8
ResNet50	72.72
EfficientNet-B0	79.3

Table 3 presents a comparative evaluation of the proposed Sequential Convolutional Neural Network (SCNN) against established deep learning architectures commonly used for facial emotion recognition. The results indicate that the proposed SCNN achieved the highest accuracy of 92.0%, surpassing VGG16 (73.28%), MobileNetV2 (71.8%), ResNet50 (72.72%), and EfficientNet-B0 (79.3%). This performance suggests that the lightweight SCNN was effective in learning discriminative facial features while maintaining a relatively simple architecture. However, since the benchmark results were obtained from different studies employing varying datasets, preprocessing techniques, training configurations, and evaluation protocols, direct comparisons should be interpreted with caution and should not be regarded as definitive evidence of the superiority of the proposed model. The comparison is presented only as a literature-based reference to provide context for the reported performance of the proposed Sequential CNN. A rigorous evaluation would require all models to be trained and tested on the same dataset using identical preprocessing procedures, hyperparameter settings, and evaluation metrics. Such controlled benchmarking will be undertaken in future work to provide a fair and comprehensive comparison with established architectures, including VGG16, MobileNetV2, ResNet50, and EfficientNet-B0.

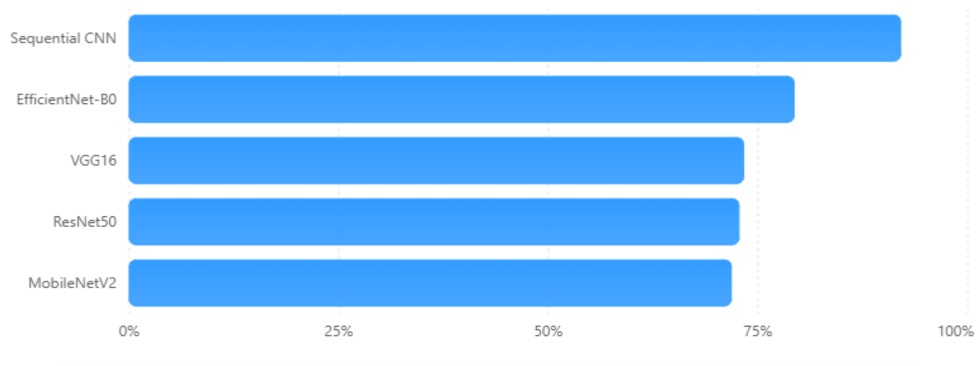

Fig. 4. Comparative Accuracy Performance of the Proposed Sequential CNN and Established CNN Architectures for Facial Emotion Recognition

Figure 4 Comparison of classification accuracy achieved by the proposed Sequential CNN and established deep learning architectures for facial emotion recognition. The proposed Sequential CNN achieved the highest reported accuracy of 92.0%, outperforming EfficientNet-B0 (79.3%), VGG16 (73.28%), ResNet50 (72.72%), and MobileNetV2 (71.8%). These results suggest that the proposed architecture demonstrated strong classification performance on the evaluated dataset. However, since the benchmark results were obtained from different studies using varying datasets and experimental protocols, the comparison should be interpreted as a literature-based reference rather than a direct experimental evaluation. The model correctly identified emotions such as *Surprised*, *Anger*, and *Happy*, showing effective feature extraction and classification capability. Although one sample of *Disgust* was misclassified as *Contempt*, the overall results indicate that the Sequential CNN achieved strong and reliable emotion prediction performance across most facial expressions.

3.2. Discussion

The experimental findings demonstrate that the proposed lightweight Sequential CNN architecture achieved strong facial emotion recognition capability while maintaining reduced computational complexity. The overall classification accuracy of 92% indicates that simplified CNN architectures can still achieve reliable predictive performance without relying on heavily parameterized deep neural networks [23]. These findings support the growing interest in lightweight affective computing frameworks suitable for deployment in embedded systems, mobile platforms, edge devices, and real-time intelligent environments. Compared with conventional deep learning architectures such as ResNet, VGGNet, and transformer-based models, the proposed Sequential CNN employs a significantly simpler architectural design while still maintaining strong classification effectiveness [20]. The reduced architectural complexity decreases computational requirements, training time, and memory consumption, making the framework more suitable for practical real-time deployment scenarios where low-latency emotion recognition and efficient resource utilization are essential. The strong classification performance achieved in this study may also be attributed to the effectiveness of the preprocessing and augmentation strategies employed during model training. Image normalization improved numerical stability and feature consistency [24]. while augmentation operations enhanced dataset diversity and improved generalization capability. These preprocessing procedures enabled the Sequential CNN model to learn robust facial representations despite variations in lighting conditions, facial orientation, and expression intensity [25]. The confusion matrix analysis further demonstrated that the majority of facial expressions were correctly classified, with only minor misclassification observed in visually similar emotional categories such as Disgust and Contempt. This limitation reflects the inherent complexity of facial emotion recognition, where subtle muscular differences may produce overlapping visual characteristics between certain emotional states.

The findings of this study contribute significantly to ongoing research in affective computing by demonstrating that lightweight CNN architectures can provide efficient and accurate facial emotion recognition suitable for practical intelligent systems. Unlike highly complex deep learning frameworks requiring extensive computational resources, the proposed lightweight model offers a more scalable and deployment-friendly solution for real-time applications [26]. The study therefore provides a transparent evaluation framework that may serve as a baseline for future investigations involving multimodal emotion recognition, real-time affective computing systems, and advanced deep learning-based facial analysis applications. Despite the strong performance achieved by the proposed Sequential CNN model, several limitations remain. First, the dataset size was relatively moderate compared with large-scale FER benchmark datasets, which may limit broader generalization capability under highly diverse real-world conditions. Although the proposed Sequential CNN achieved an accuracy of 92% on the evaluation subset, direct comparison with results reported in the literature should be interpreted with caution. Previous studies on facial emotion recognition have reported accuracies ranging from approximately 71% to 79% for transfer-learning-based architectures such as MobileNetV2, ResNet50, VGG16, and EfficientNet-B0 on benchmark FER datasets. These models benefit from substantially deeper architectures and pretrained feature representations that improve their ability to capture complex facial patterns. However, the reported results were obtained under different experimental settings, datasets, and evaluation protocols. Consequently, the present study cannot claim superiority over these architectures without conducting controlled experiments using the same dataset and training conditions.

Future work will therefore include a direct empirical comparison against VGG16, MobileNetV2, ResNet50, and EfficientNet-based models to establish the relative strengths and limitations of the proposed Sequential CNN architecture [27]. Second, the study utilized static facial images rather than video-based temporal emotion sequences, potentially limiting the model's ability to capture dynamic facial transitions and temporal emotional variations. Additionally, the proposed framework focused exclusively on facial emotion recognition and should not be interpreted as a clinically validated depression detection system because the employed dataset contained emotion labels rather than psychiatric diagnostic annotations [28]. Future research should therefore investigate larger and more diverse datasets, multimodal affective computing systems integrating facial expressions, speech patterns, physiological signals, and behavioral data, as well as transfer learning and attention-based architectures capable of learning finer-grained facial representations [29][30] Furthermore, the integration of transformer-based networks and explainable artificial intelligence techniques may improve recognition robustness, interpretability, and generalization capability. The deployment of lightweight FER systems in real-time mobile, embedded, and edge computing environments additionally represents an important direction for future practical implementation and optimization.

4. CONCLUSION

This study developed a lightweight Sequential Convolutional Neural Network (S-CNN) for multi-class facial emotion recognition and demonstrated its effectiveness in accurately classifying facial expressions across eight emotional categories. The proposed model achieved strong performance, attaining 92% accuracy, 87% weighted precision, 92% weighted recall, and 89% weighted F1-score, while maintaining low computational complexity suitable for real-time applications. The findings indicate that lightweight CNN architectures can provide efficient and reliable facial emotion recognition without requiring highly complex deep learning frameworks. Although the study was limited to a single dataset and literature-based comparisons with benchmark models, the results highlight the potential of the proposed approach for affective computing applications. Future research should focus on larger datasets, direct benchmarking against state-of-the-art architectures, statistical validation, and real-world deployment to further improve robustness and generalizability.

Declarations

Author Contribution

Godfrey Perfectson OISE: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources; Immunhierokene Clinton OBRORINDO: Formal analysis, Investigation, Resources, Data Curation, Writing – Original Draft; Roli Lydia OSHASHA: Writing – Review & Editing, Visualization, Supervision, Project administration; Kevin Chinedu PIUS: Supervision, Project administration, Funding acquisition; Felix Oshiorenoya ULOKO: Validation, Data Curation, Writing – Re-view & Editing, Visualization.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

The data presented in this study are available in the manuscript.

Acknowledgment

Not applicable.

Conflicts of Interest

The author declares no conflicts of interest

REFERENCES

[1] Z. Jiang, S. Harati, A. Crowell, H. S. Mayberg, S. Nemati, and G. D. Clifford, “Classifying Major Depressive Disorder and Response to Deep Brain Stimulation Over Time by Analyzing Facial Expressions,” IEEE Trans. Biomed. Eng., vol. 68, no. 2, pp. 664–672, Feb. 2021, doi: 10.1109/TBME.2020.3010472.

[2] N. B. Krishna, R. Y. Kumar, M. I. Bhanu, P. O. Reddy, R. Anjali, and P. R. Reddy, “Exploring Machine Learning Algorithms for the Detection of Depression,” in 2024 International Conference on Smart Systems for applications in Electrical Sciences (ICSSES), Tumakuru, India: IEEE, May 2024, pp. 1–6, doi: 10.1109/ICSSES62373.2024.10561447.

[3] P. Vinayagam, P. Kishore Kumar, and M. Kandhan, “FPGA-Based Automated EEG Analysis for Efficient Emotion Detection,” in 2024 Third International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE), Ballari, India: IEEE, Apr. 2024, pp. 1–6, doi: 10.1109/ICDCECE60827.2024.10548113.

[4] H. Zhang, T. Zuo, Z. Chen, X. Wang, and P. Z. H. Sun, “Evolutionary Ensemble Learning for EEG-Based Cross-Subject Emotion Recognition,” IEEE J. Biomed. Health Inform., vol. 28, no. 7, pp. 3872–3881, July 2024, doi: 10.1109/JBHI.2024.3384816.

[5] Z. Jiang et al., “Multimodal Mental Health Digital Biomarker Analysis From Remote Interviews Using Facial, Vocal, Linguistic, and Cardiovascular Patterns,” IEEE J. Biomed. Health Inform., vol. 28, no. 3, pp. 1680–1691, Mar. 2024, 10.1109/JBHI.2024.3352075.

[6] D. Yang et al., “Using Wearable and Structured Emotion-Sensing-Graphs for Assessment of Depressive Symptoms in Patients Undergoing Treatment,” IEEE Sensors J., vol. 24, no. 3, pp. 3637–3648, Feb. 2024, doi: 10.1109/JSEN.2023.3339498.

[7] H. V, S. T, and M. Siddappa, “Facial Emotion Recognition Method Based on Canny Edge Detection Using Convolutional Neural Network,” in 2023 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India: IEEE, Nov. 2023, pp. 425–430, doi: 10.1109/ICCCIS60361.2023.10425079.

[8] D.-H. Lee and J.-H. Yoo, “CNN Learning Strategy for Recognizing Facial Expressions,” IEEE Access, vol. 11, pp. 70865–70872, 2023, doi: 10.1109/ACCESS.2023.3294099.

[9] G. D. Jadhav, S. D. Babar, and P. N. Mahalle, “Exploratory Data Analysis and Algorithmic Comparison for Early Depression Detection,” in 2024 IEEE 4th International Conference on ICT in Business Industry & Government (ICTBIG), Indore, India: IEEE, Dec. 2024, pp. 1–7, doi: 10.1109/ICTBIG64922.2024.10911717.

[10] A. Sharma, A. Saxena, A. Kumar, and D. Singh, “Depression Detection Using Multimodal Analysis with Chatbot Support,” in 2024 2nd International Conference on Disruptive Technologies (ICDT), Greater Noida, India: IEEE, Mar. 2024, pp. 328–334, doi: 10.1109/ICDT61202.2024.10489080.

[11] C. Białek, A. Matiolański, and M. Grega, “An Efficient Approach to Face Emotion Recognition with Convolutional Neural Networks,” *Electronics*, vol. 12, no. 12, p. 2707, Jun. 2023, doi: 10.3390/electronics12122707.

[12] D. D. Mandave and L. V. Patil, “A Statistically Rigorous Comparison of MobileNetV2 and EfficientNet-B0 for Facial Expression Recognition on the FER2013 Benchmark,” *Inf. Dyn. Appl.*, vol. 4, no. 4, pp. 189–200, Dec. 2025, doi: 10.56578/ida040401.

[13] Y. Khaireddin and Z. Chen, “Facial Emotion Recognition: State of the Art Performance on FER2013,” 2021, *arXiv*. doi: 10.48550/ARXIV.2105.03588.

[14] T. Lee et al., “A Deep Learning Driven Simulation Analysis of the Emotional Profiles of Depression Based on Facial Expression Dynamics,” Clin Psychopharmacol Neurosci, vol. 22, no. 1, pp. 87–94, Feb. 2024, doi: 10.9758/cpn.23.1059.

[15] V. S. Bakkialakshmi and T. Sudalaimuthu, “A Survey on Affective Computing for Psychological Emotion Recognition,” in 2021 5th International Conference on Electrical, Electronics, Communication, Computer

Technologies and Optimization Techniques (ICEECCOT), Mysuru, India: IEEE, Dec. 2021, pp. 480–486, doi: 10.1109/ICEECCOT52851.2021.9707947.
[16] M. Yang et al., “LMS-VDR: Integrating Landmarks into Multi-scale Hybrid Net for Video-Based Depression Recognition,” in Pattern Recognition and Computer Vision, vol. 15040, Z. Lin, M.-M. Cheng, R. He, K. Ubul, W. Silamu, H. Zha, J. Zhou, and C.-L. Liu, Eds., in Lecture Notes in Computer Science, vol. 15040, Singapore: Springer Nature Singapore, 2025, pp. 299–312, doi: 10.1007/978-981-97-8792-0_21.
[17] M. Senthil Sivakumar, T. Gurumekala, L. Megalan Leo, and R. Thandaiah Prabu, “Expert System for Smart Virtual Facial Emotion Detection Using Convolutional Neural Network,” Wireless Pers Commun, vol. 133, no. 4, pp. 2297–2319, Dec. 2023, doi: 10.1007/s11277-024-10867-0.
[18] C. Gupta and V. Khullar, “Exploring the Role of Conventional and Non-Conventional Technologies in Identifying MDD,” in 2022 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI), Delhi, India: IEEE, Dec. 2022, pp. 1–6, doi: 10.1109/SOLI57430.2022.10294421.
[19] Y. Guo, L. Wang, Y. Xiao, and Y. Lin, “A Personalized Spatial-Temporal Cold Pain Intensity Estimation Model Based on Facial Expression,” IEEE J. Transl. Eng. Health Med., vol. 9, pp. 1–8, 2021, doi: 10.1109/JTEHM.2021.3116867.
[20] X. Li, W. Guo, and H. Yang, “Depression severity prediction from facial expression based on the DRR_DepressionNet network,” in 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Seoul, Korea (South): IEEE, Dec. 2020, pp. 2757–2764, doi: 10.1109/BIBM49941.2020.9313597.
[21] Q. Li et al., “Machine Learning-Based Prediction of Depressive Disorders via Various Data Modalities: A Survey,” IEEE/CAA J. Autom. Sinica, vol. 12, no. 7, pp. 1320–1349, July 2025, doi: 10.1109/JAS.2025.125393.
[22] S. Zhang et al., “MTDAN: A Lightweight Multi-Scale Temporal Difference Attention Networks for Automated Video Depression Detection,” IEEE Trans. Affective Comput., vol. 15, no. 3, pp. 1078–1089, July 2024, doi: 10.1109/TAFFC.2023.3312263.
[23] M. Li, Z. Lu, Q. Cao, J. Gao, and B. Hu, “Automatic Assessment Method and Device for Depression Symptom Severity Based on Emotional Facial Expression and Pupil-Wave,” IEEE Trans. Instrum. Meas., vol. 73, pp. 1–15, 2024, doi: 10.1109/TIM.2024.3415778.
[24] M. Niu, Y. Li, J. Tao, X. Zhou, and B. W. Schuller, “DepressionMLP: A Multi-Layer Perceptron Architecture for Automatic Depression Level Prediction via Facial Keypoints and Action Units,” IEEE Trans. Circuits Syst. Video Technol., vol. 34, no. 9, pp. 8924–8938, Sept. 2024, doi: 10.1109/TCSVT.2024.3382334.
[25] J. Ye, Y. Yu, Y. Zheng, Y. Liu, and Q. Wang, “Dep-FER: Facial Expression Recognition in Depressed Patients Based on Voluntary Facial Expression Mimicry,” IEEE Trans. Affective Comput., vol. 15, no. 3, pp. 1725–1738, July 2024, doi: 10.1109/TAFFC.2024.3370103.
[26] G. Bargshady and R. Goecke, “Estimating Depression Severity from Long-Sequence Face Videos via an Ensemble Global Diverse Convolutional Model,” in 2023 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Port Macquarie, Australia: IEEE, Nov. 2023, pp. 296–303, doi: 10.1109/DICTA60407.2023.00048.
[27] G. Bargshady and R. Goecke, “Estimating Depression Severity from Long-Sequence Face Videos via an Ensemble Global Diverse Convolutional Model,” in 2023 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Port Macquarie, Australia: IEEE, Nov. 2023, pp. 296–303, doi: 10.1109/DICTA60407.2023.00048.
[28] L. He et al., “LMTformer: facial depression recognition with lightweight multi-scale transformer from videos,” Appl Intell, vol. 55, no. 3, p. 195, Feb. 2025, doi: 10.1007/s10489-024-05908-x.
[29] M. Kolahdouzi, A. Sepas-Moghaddam, and A. Etemad, “FaceTopoNet: Facial Expression Recognition Using Face Topology Learning,” IEEE Trans. Artif. Intell., vol. 4, no. 6, pp. 1526–1539, Dec. 2023. https://doi.org/10.1109/TAI.2022.3207450.
[30] G. P. OISE, P. O. EJENARHOME, and A. S. OYEDOTUN, “Enhancing road safety through deep learning-based drowsiness detection using vision AI,” *Journal of Electrical Systems and Information Technology*, vol. 13, no. 1, p. 30, Mar. 2026, doi: 10.1186/s43067-026-00335-z.