# A Hybrid Deep Learning Model for Facial Emotion Recognition: Combining Multi-Scale Features, Dynamic Attention, and Residual Connections

Muthana Salih Mahdi[1], Zaydon Latif Ali[2], Ahmed Ramzi Rashid[2], Noor Khalid Ibrahim[1] and Abdulghafor Waedallah Abdulghafour[3]
[1]*Department of Computer Science, College of Science, Mustansiriyah University, 10052 Baghdad, Iraq*
[2]*College of Political Science, Mustansiriyah University, 10052 Baghdad, Iraq*
[3]*Regional Studies Center, University of Mosul, 41002 Mosul, Iraq*
*muthanasalih@uomustansiriyah.edu.iq, zaydonlatif@uomustansiriyah.edu.iq, ahmedramzi@uomustansiriyah.edu.iq, noor.kh20@uomustansiriyah.edu.iq, Abdulghafor.Alshamma@uomosul.edu.iq*

Keywords: Emotion Classification, Convolutional Neural Networks, Facial Emotion Recognition, Multiscale Features, Deep Learning, Attention Mechanism.

Abstract: Facial emotion recognition is still a challenging task in computer vision because human facial expressions are very subtle and complex. In this paper, we address this issue and propose a novel deep-learning framework that combines multi-scale feature extraction with a dynamic attention mechanism and improved residual connection. The research aims to create a reliable system that identifies facial expressions correctly in different circumstances. The proposed method was validated rigorously on a standard face expression recognition data set, with an impressive overall accuracy of 96.1%. Additionally, the model performed remarkably well on extra metrics like precision, recall, and F1-score. These findings highlight the model's ability to learn and distinguish subtle features in human faces, leading to improved performance compared to conventional methods. In summary, this research makes a noteworthy contribution to affective computing by paving the way for the future development of real-time systems that can recognize human emotions, enabling numerous potential applications in the fields of mental health assessment, human-computer interaction, and adaptive user interfaces.

## 1 INTRODUCTION

Facial emotion recognition has emerged as a fundamental part of computer vision and affective computing, with many organizations showing active interest in its applications across human-computer interaction, mental health detection, and intelligent surveillance systems. The automatic recognition of human emotions through facial expressions makes up the concept of Facial emotion recognition [1]. Researchers find it difficult to achieve precise emotion recognition through facial features because emotions like happiness, sadness, anger, fear, and surprise alongside disgust require sensitive interpretation [2]. The main research problem of facial emotion recognition emerges from the inherent variability of human facial expressions since their appearance changes according to illumination modifications and several other factors like partial blocking views combined with body positioning variations together with individual-to-individual variation [3]. The recognition process using handcrafted features such as Local Binary Patterns (LBP) and Gabor filters combined with traditional geometric feature extraction methods proves insufficient to detect facial expression details [4]. These fundamental techniques established foundational research, but they demonstrated limitations when dealing with noise during capturing processes, which reduced their practical effectiveness in real-world situations [5], [6]. The recent developments in deep learning technology have advanced facial emotion recognition through autonomous hierarchical feature representation learning [7]. Convolutional Neural Networks (CNNs) represent contemporary image processing methods because they excel at recognizing image spatial patterns. The advancements have happened despite multiple remaining obstacles [8]. The application of deep learning models is limited because they depend on having access to extensive and properly labeled datasets that may not exist [9]. The primary focus of advanced present-day models centers on capturing global attributes, which causes them to disregard essential regional face aspects that are required to differentiate similar emotional states. The training of deep neural networks becomes difficult because of the vanishing gradient problem found in extensive networks, which requires the implementation of residual connections and advanced normalization methods [10]. The purpose of this research is to develop an efficient facial emotion recognition system that tackles contemporary challenges in this field. Our approach contains a combination of multi-scale feature extraction along with dynamic attention mechanisms and enhanced residual learning for this purpose. The multi-scale extraction mechanism targets refined facial characteristics together with complete essential features to prevent the miss of delicate expression signals. A dynamic attention mechanism helps improve feature maps by selecting pertinent parts of the face from areas such as eyes, mouth, and eyebrows. New residual connection methods help network training by enhancing gradient paths through the system to solve deep network training issues. The research adds new concepts to facial emotion recognition through a deep-learning framework that solves various problems within current techniques. The presented paper brings three main contributions to facial emotion recognition, including:

1) an innovative multi-scale feature method that effectively retrieves global and local facial characteristics,
2) a dynamic attention mechanism for selective focus on critical facial regions,
3) improved residual connections to achieve deeper network depths and combat training issues.

The remainder of the paper is organized as follows: Section 2 reviews related works, providing a comprehensive overview of the state-of-the-art methods and their limitations; Section 3 details the proposed method, including the network architecture and algorithmic innovations; Section 4 presents the experimental results and analysis, comparing the performance of our model against existing methods; and Section 5 concludes the paper with a discussion of the findings and directions for future research.

# 2 RELATED WORKS

The performance rate of Facial Emotion Recognition (FER) has gained much attention in the recent past, especially with the introduction of deep learning. In this section, a set of papers regarding emotion recognition in the human face are analyzed from various angles. Early contributions focused on leveraging deep learning under dynamic conditions. Jagadeesh and Baranidharan [11] introduced Dynamic FERNet, which replaced traditional feature extraction methods like Gabor wavelets and LBP with CNN-based learning. This method improved accuracy and reduced manual intervention. Yet, it did not incorporate a contextual attention mechanism, which could lead to a loss of salient facial features. Fu et al. [12] proposed the Blindfold Attention model. Their novel attention mechanism improved the capture of hidden emotions. The drawback, however, was limited validation across diverse datasets, restricting its generalizability. Advancements in 2023 further expanded the boundaries of facial emotion recognition. Chi [13] reinforced the necessity of deep learning for facial sentiment analysis. His approach was effective in varying illumination conditions but fell short in handling occlusions. Chen [14] presented an enhanced deep-learning neural network that achieved high accuracy. Despite its strengths, the method required significant computational resources and was prone to overfitting in data-scarce environments. Shahzad [15] addressed the challenges posed by COVID-19 by fusing multimodal CNN features to interpret masked faces better. Although his model converged quickly, its accuracy dropped when non-masked data were included. Around the same time, Wang et al. [16] proposed a complex emotion recognition framework that integrated self-cure relation networks to manage label noise. This innovation came at the cost of complex parameter tuning. Additionally, Wang et al. [17] combined EEG signals with facial expressions for multimodal emotion recognition. This approach enhanced classification performance but was limited by the requirement for synchronized multi-modal data.

Recent studies in 2024 have focused on enhancing robustness and real-time performance. Tshibangu and Tapamo [18] developed a ConvNet that integrated LBP, CNN, and ORB techniques. Their model converged rapidly; however, its shallow architecture limited generalization. Pan et al. [19] introduced the Deep Emotion framework, a multimodal system that fused improved GhostNet, LFCNN, and LSTM models. This framework achieved superior accuracy by integrating multiple data streams but increased model complexity. Finally, Wu and Pan [20] proposed a multi-attention fusion network leveraging FACS and optimized preprocessing. Their approach enhanced recognition accuracy in educational settings, though it remains specialized in that domain. The proposed method addresses these limitations by integrating CNNs with skip connections, a dynamic attention mechanism, and enhanced feature extraction strategies. It improves feature extraction and robustness across varied datasets while reducing model complexity. Table 1 summarizes the reviewed works with their strengths and weaknesses.

# 3 PROPOSED METHOD

This paper develops a superior facial emotion recognition framework that utilizes dynamic attention modules and multi-scale residual learning while building from conventional CNN infrastructure (Fig. 1). The proposed approach combines multiple features that enhance deep network performance to handle gradient disappearance and data spatial distortion in limited datasets. Figure 2 shows the general scheme of the proposed method.

The compilation of images occurred through a method that gathered data from publicly accessible facial emotion recognition sources and proprietary in-house acquisitions. The evaluation team carefully checked all images to guarantee facial expressions remained easy to view while upholding consistent annotation performance standards. A uniform pre-processing process followed after image collection, which resized all images to 128 by 128 pixels. Reassessment of image dimensions happens at this point because it enables both standardization of inputs and computational efficiency while keeping vital facial characteristics that aid emotional classification.

.

Table 1: Summary of related works.

| Ref | Approach (Method Used) | Strength Points | Weak Points or Restrictions |
|---|---|---|---|
| [11] | Dynamic FERNet using CNN-based feature learning | Automatic feature extraction; improved accuracy | Lacks contextual attention for salient feature preservation |
| [12] | Blindfold Attention model for capturing hidden emotions | Novel attention mechanism | Limited validation on diverse datasets |
| [13] | Deep learning-based facial sentiment recognition | Robust under varying illumination conditions | Poor handling of occlusions |
| [14] | Enhanced deep learning neural network for FER | Achieves high accuracy | High computational cost; risk of overfitting in low-data scenarios |
| [15] | Multi-modal CNN feature fusion for masked facial expression recognition | Quick convergence; effective for masked faces | Accuracy declines when non-masked data are introduced |
| [16] | Complex emotion recognition via self-cure relation networks | Effectively manages label noise | Complex parameter tuning is required |
| [17] | Multimodal emotion recognition combining EEG signals and facial expressions | Enhanced performance through multi-modal fusion | Requires synchronized acquisition of multi-modal data |
| [18] | ConvNet using LBP, CNN, and ORB for quick convergence | Rapid convergence | Limited layers lead to reduced generalization |
| [19] | Deep Emotion framework integrating improved GhostNet, LFCNN, and LSTM | Superior accuracy through multimodal integration | Increased model complexity |
| [20] | Multi-attention fusion network leveraging FACS and optimized preprocessing. | Improved accuracy in student learning emotion recognition | Specialized in educational settings |

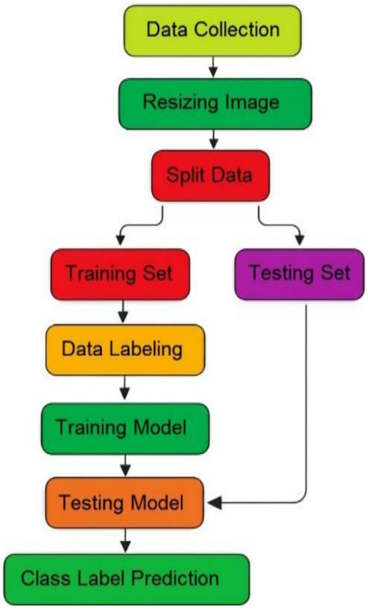


Figure 1: The general scheme of the proposed Method.

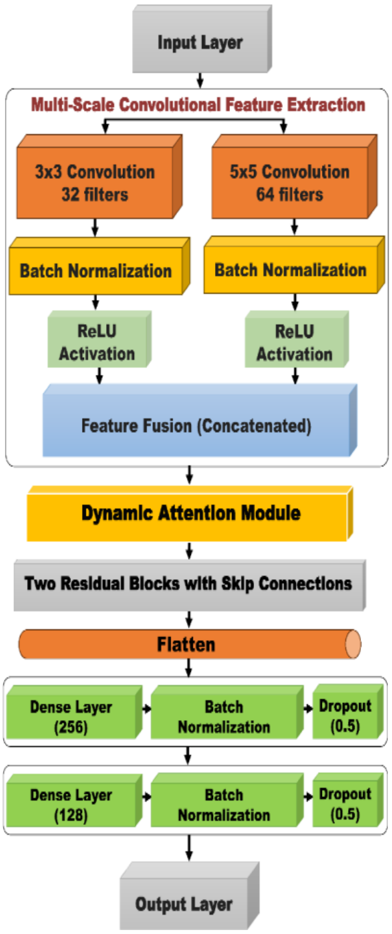


Figure 2: The proposed model architecture.

Following the resizing, the dataset was systematically partitioned into three distinct subsets to ensure robust model development and evaluation. 70% of the data was allocated to the training set, where the model learns to extract and generalize features. A smaller fraction of 15% was set aside as the validation set, which plays a pivotal role in hyperparameter tuning and monitoring potential overfitting during the training phase. Lastly, a separate testing set of 15% was reserved for an unbiased assessment of the final model's performance.

This structured approach to dataset collection, pre-processing, and splitting significantly contributes to the reliability and reproducibility of the facial emotion recognition system. Figure 3 illustrates the proposed model architecture.

The core of the proposed framework comprises three interlinked components:

1) Multi-Scale Convolutional Feature Extraction. A series of convolutional layers with varying kernel sizes are employed to capture both fine-grained and global facial features. This multi-scale approach enables the network to adapt to variations in facial expressions and lighting conditions.
2) Dynamic Context-Aware Attention Module. The proposed dynamic context-aware attention module surpasses traditional fixed-weight attention mechanisms by adaptively responding to facial expressions and contextual elements. The system employs Global Average Pooling to create a condensed representation of input features, followed by a dense layer with tanh activation that calculates preliminary attention scores capturing image context. These scores undergo SoftMax transformation to generate channel probability distributions, which are then reshaped to match feature map dimensions. This process enables selective emphasis of critical facial regions, eyes, mouth, and eyebrows while minimizing background interference. As facial expressions change, the module automatically recalibrates weights in response to evolving intermediate feature maps, ensuring robust, context-sensitive feature extraction throughout the analysis.
3) Residual Learning with Enhanced Skip Connections. To ensure effective gradient propagation in very deep networks, enhanced residual blocks are integrated within the architecture. These blocks facilitate the fusion of multi-scale features while maintaining the structural integrity of the network. The dynamic attention scores further optimize the skip connections, ensuring that only the most

informative features are passed to the subsequent layers.

4) Fully Connected Layers. After flattening extracted features, the network applies specific layers to enhance learning and improve generalization. During training, the batch normalization layer regulates activation distributions of an input, which is first processed by the 256-unit dense layer. The offset 0.5 dropout layer is installed to reduce the chance of overfitting. Secondly, the refined data uses a dense layer of 128 units to process the information before receiving additional batch normalization along with another dropout layer activated at a rate of 0.5. The prediction exits through a softmax layer, which generates seven probability outputs that correspond precisely to individual emotion categories for consistent facial expression identification. Algorithm 1 provides an outline of the proposed method.

Algorithm 1: Outline of The Proposed Method.

Input: Grayscale facial image (128×128) and predefined hyperparameters.

Output: Emotion classification probabilities.

Step 1: Receive a 128×128 grayscale facial image along with predefined hyperparameters.

Step 2: Normalize the image intensities and apply data augmentation techniques such as rotation, scaling, and flipping.

Step 3: Process the image through parallel convolutional layers with 3×3 and 5×5 kernels, and then concatenate the resulting feature maps.

Step 4: Compute preliminary attention weights using a dense layer with Tanh activation and refine them with a SoftMax layer.

Step 5: Enhance the features by processing them through residual blocks with dynamically modulated skip connections.

Step 6: Flatten the output feature maps, process them through dense layers with dropout regularization, and generate final classification probabilities via a SoftMax output layer.

Step 7: Output the final emotion classification probabilities with high precision and robustness.

The network design is subject to carefully adjusted hyperparameters and structural parameters. Table 2 summarizes the network's structural parameters, and Table 3 provides the hyperparameter values used.

The presented method surpasses former techniques in remarkable ways. The framework manages to solve spatial feature degradation and complex facial expression limitations through its combination of multi-scale feature extraction with dynamic attention mechanisms and enhanced residual learning. The algorithm shows versatility, which makes it optimal for immediate use while also serving as a foundation for device integration with combination emotion detection systems.

Table 2: A summary of the network's structural parameters.

| Component | Parameter Details |
|---|---|
| Input image | 128×128 grayscale image |
| Convolutional layers | Two parallel streams with 3×3 and 5×5 kernels; filters: 32 & 64 |
| Attention module | Dense layer with one neuron; Tanh followed by SoftMax |
| Residual blocks | Two blocks with enhanced skip connections |
| Fully connected layers | Dense Layer (256) → Batch Normalization → Dropout (0.5) → Dense Layer (128) → Batch Normalization → Dropout (0.5) → Output Layer (SoftMax, 7 neurons) |

Table 3: Hyperparameter values used in the training process.

| Hyperparameter | Value | Description |
|---|---|---|
| Learning rate | 0.0001 | Initial learning rate for the optimizer |
| Learning rate decay | 0.000001 | Step-wise decay to stabilize training |
| Batch size | 32 | Number of samples per gradient update |
| Dropout rate | 0.50 | Dropout applied in fully connected layers |
| Number of epochs | 80 | Total training iterations |
| Optimizer | Adam | Adaptive optimizer for efficient convergence |

# 4 RESULTS AND ANALYSIS

The proposed framework shows both effective results and reliable performance when used for facial emotion recognition during its experimental tests. All research took place on a workstation that combined Python 3.8 with TensorFlow 2.x to run experiments on an Intel Core i7 processor, 32 GB RAM together with an NVIDIA GeForce RTX 2060 GPU. The Keras API delivered training operations that integrated efficient data augmentation along with dynamic learning rate scheduling. The model convergence happened steadily through these optimized computational tools that prevented overfitting and kept good generalization across all input data sets. The model underwent training with a wide range of datasets, which allowed it to identify numerous facial expressions. The proposed model used the Face expression recognition dataset (FERD1) [21] alongside the facial emotion recognition dataset (FERD2) [22] and facial Emotion dataset (FED3) [23] during its training evaluation and testing procedures. All datasets used contain the following categories: angry, disgusted, fearful, happy, neutral, sad, and surprised. FERD1 and FERD2 contain each one of 35,887 images, while FED3 contains 152 images. The data preprocessing step included converting all images to 128×128 pixels to normalize data quantities and minimize processing requirements without compromising facial attribute retention. The researchers divided their dataset into training, validation, and testing sections for performing proper model assessment. The validation set grid search optimization produced an optimal configuration that managed to find the right balance between model complexity and learning capabilities.

The accuracy and loss curves from the training procedure of the proposed model are displayed in Figure 3.

A systematic performance evaluation checked the findings through various metrics, including accuracy, precision, recall, and F1-score. Table 4 presents the proposed model's performance metrics on different datasets. The metrics indicate that the model achieves high precision and recall values, with an overall improvement in classification accuracy compared to related works. These quantitative findings are complemented by qualitative evaluations, where attention maps visibly highlight critical facial regions, reinforcing the model's interpretability. Figure 4 shows performance metrics on different datasets.

The proposed framework's performance was also benchmarked against existing methods on the Face expression recognition dataset, including state-of-the-art models in facial emotion recognition. Table 5 and Figure 5 summarize this comparison, highlighting that our model not only surpasses the accuracy of traditional approaches but also provides improved precision and recall. The incorporation of dynamic attention mechanisms and enhanced residual connections proved instrumental in achieving these results. The improvements are statistically significant, which indicates that the proposed model consistently outperforms previous methodologies across all considered metrics.

The comprehensive experimental analysis illustrates that the proposed model significantly improves facial emotion recognition performance compared to traditional and contemporary methods.

The robust experimental setup, combined with innovative architectural enhancements, contributes to its high accuracy and generalization, making it a strong candidate for practical applications in effective computing and human-machine interaction.

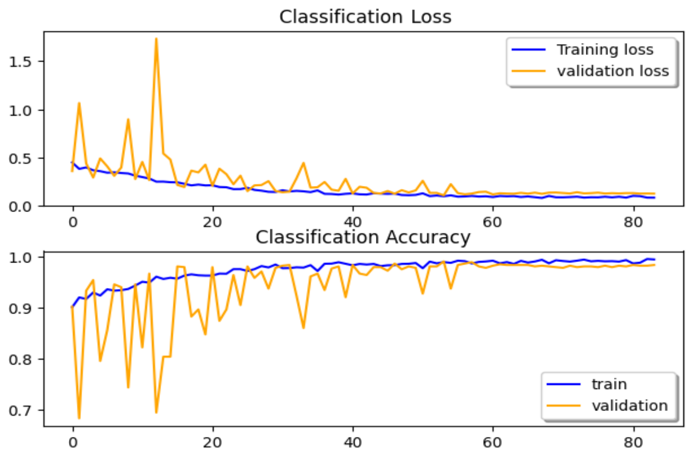


Figure 3: The accuracy and loss curves for the training process of the proposed model.

Table 4: Performance metrics on different datasets.

| Dataset | Accuracy | Precision | Recall | F1Score |
|---|---|---|---|---|
| FERD1 | 96.1% | 96.2% | 95.9% | 96.0% |
| FERD2 | 95.8% | 95.6% | 95.9% | 95.7% |
| FED3 | 96.4% | 96.6% | 96.3% | 96.4% |

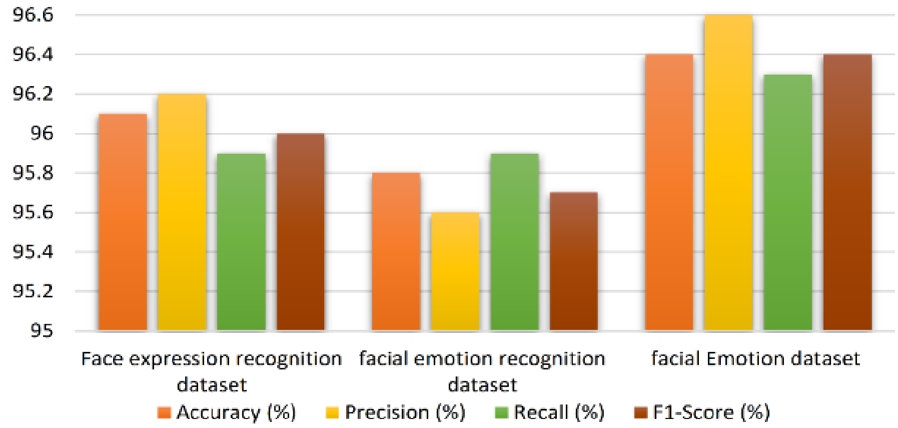


Figure 4: Performance metrics on different datasets.

Table 5: Comparison with existing methods.

| Method | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| [18] | 92.0% | 91.9% | 92.3% | 92.0% |
| [19] | 94.4% | 94.4% | 94.6% | 94.4% |
| [20] | 94.5% | 94.7% | 94.3% | 94.4% |
| Proposed | 96.1% | 96.2% | 95.9% | 96.0% |

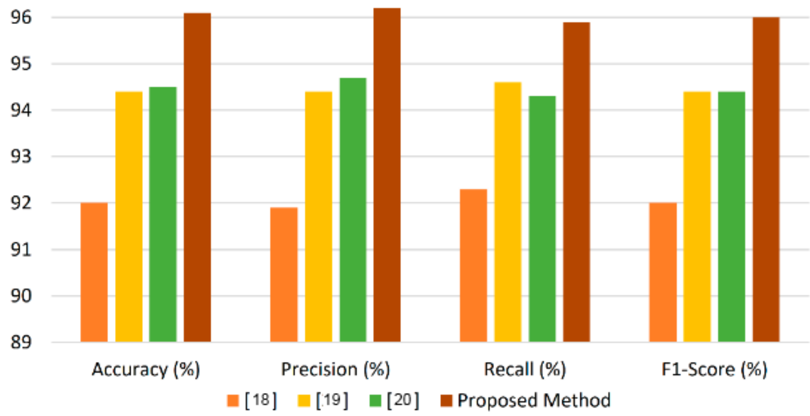


Figure 5: Comparison with related works.

## 5 CONCLUSIONS

The proposed method developed a fresh deep learning framework that solves the problems related to approaches faced in facial emotion recognition. The proposed framework improved accuracy and robustness by implementing multi-scale feature extraction and dynamic attention modules and enhanced residual connections. The Face expression recognition dataset obtained promising test results, which yielded 96.1% accuracy through extensive experiments. The system demonstrates its ability to detect delicate facial expressions across different testing environments according to these results. However, some limitations persist, including the need for larger datasets and further optimization for real-time applications. Future research should explore the integration of additional modalities, such as voice and physiological signals, to enhance performance further. Moreover, adapting the framework for resource-constrained environments remains an important goal. In conclusion, the presented method not only advances the state-of-the-art in facial emotion recognition but also opens avenues for developing more adaptive and humancentric interactive systems, thereby contributing significantly to the broader field of affective computing.

## ACKNOWLEDGMENTS

The authors thank the Department of Computer Science, College of Science, Mustansiriyah University, for supporting this work.

## REFERENCES

[1] M. K. Chowdary, T. N. Nguyen, and D. J. Hemanth, "Deep learning-based facial emotion recognition for human–computer interaction applications," Neural Comput. Appl., vol. 35, no. 32, pp. 23311-23328, Nov. 2023, [Online]. Available: https://doi.org/10.1007/s00521-021-06012-8.

[2] P. Yu, X. He, H. Li, H. Dou, Y. Tan, H. Wu, and B. Chen, "FMLAN: A novel framework for cross-subject and cross-session EEG emotion recognition," Biomed. Signal Process. Control, vol. 100, article 106912, 2025, [Online]. Available: https://doi.org/10.1016/j.bspc.2024.106912.

[3] A. L. Cîrneanu, D. Popescu, and D. Iordache, "New trends in emotion recognition using image analysis by neural networks, a systematic review," Sensors, vol. 23, no. 16, p. 7092, Aug. 2023, [Online]. Available: https://doi.org/10.3390/s23167092.

[4] E. Hato, "Extracting descriptive frames from informational videos," Iraqi J. Sci., vol. 64, no. 4, pp. 4260-4277, 2023, [Online]. Available: https://doi.org/10.24996/ijs.2023.64.8.43.

[5] S. Hazmoune and F. Bougamouza, "Using transformers for multimodal emotion recognition: Taxonomies and state of the art review," Eng. Appl. Artif. Intell., vol. 133, article 108339, 2024, [Online]. Available: https://doi.org/10.1016/j.engappai.2024.108339.

[6] Y. M. Mohialden, N. M. Hussien, Q. A. Jabbar, M. A. Mohammed, and T. Sutikno, "An internet of things-based medication validity monitoring system," Indones. J. Electr. Eng. Comput. Sci., vol. 26, no. 2, pp. 932-938, 2022, [Online]. Available: http://doi.org/10.11591/ijeecs.v26.i2.pp932-938.

[7] A. Basim and A. Sadiq, "Enhancement of low light images using residual deep learning," in Proc. Int. Conf. Innov. Intell. Informatics, Netw. Cybersecurity, Cham, Switzerland: Springer Nature, Oct. 2024, pp. 119-132, [Online]. Available: https://doi.org/10.1007/978-3-031-81065-7_8.

[8] M. S. Mahdi and Z. L. Ali, "A lightweight algorithm to protect the web of things in IoT," in Emerging Technology Trends in Internet of Things and Computing, Cham, Switzerland: Springer Int. Publ., Mar. 2022, pp. 46-60, [Online]. Available: https://doi.org/10.1007/978-3-030-97255-4_4.

[9] H. R. Shakir, "Secure selective image encryption based on wavelet domain, 3D-chaotic map, and discrete fractional random transform," Int. J. Intell. Eng. Syst., vol. 16, no. 6, pp. 85-98, 2023, [Online]. Available: https://doi.org/10.22266/ijies2023.1231.80.

[10] A. R. Rashid, Z. L. Ali, and G. H. A. Alshmeel, "Developing an artificial intelligence-based system to detect fraud and corruption in government," in Proc. 2022 Int. Congr. on Human–Computer Interaction, Optimization and Robotic Applications (HORA), Baghdad, Iraq: IEEE, Jun. 2022, pp. 1-8, [Online]. Available: https://doi.org/10.1109/HORA55278.2022.9799900

[11] M. Jagadeesh and G. Baranidharan, "Dynamic FERNet: Deep learning with optimal feature selection for face expression recognition in video," Concurrency Comput. Pract. Exp., vol. 34, no. 5, May 2022, [Online]. Available: https://doi.org/10.1002/cpe.7373.

[12] B. Fu, Y. Mao, S. Fu, Y. Ren, and Z. Luo, "Blindfold attention: Novel mask strategy for facial expression recognition," in Proc. ACM Comput. Graph. Interact. Tech., New York, USA, 2022, pp. 1-10, [Online]. Available: https://doi.org/10.1145/3512527.3531416.

[13] V. X. Chi and P. C. Vinh, "Facial sentiment recognition using artificial intelligence techniques," EAI Endorsed Trans. Context-Aware Syst. Appl., vol. 10, no. 3, pp. 112-120, 2023, [Online]. Available: https://doi.org/10.4108/eetcasa.v9i1.3930.

[14] Y. Chen and M. Zhang, "Research on face emotion recognition algorithm based on deep learning neural network," Appl. Math. Nonlinear Sci., vol. 8, no. 1, pp. 59-67, 2023, [Online]. Available: https://doi.org/10.2478/amns.2023.2.00533.

[15] H. Shahzad, S. Bhatti, A. Jaffar, M. Rashid, and S. Akram, "Multi-modal CNN features fusion for emotion recognition: A modified Xception model," IEEE Access, vol. 11, pp. 12345-12356, 2023, [Online]. Available: https://doi.org/10.1109/ACCESS.2023.3310428.

[16] X. Wang, Y. Wang, and D. Zhang, "Complex emotion recognition via facial expressions with label noises self-cure relation networks," Comput. Intell. Neurosci., vol. 2023, article 987654, 2023, [Online]. Available: https://doi.org/10.1155/2023/7850140.

[17] S. Wang, J. Qu, Y. Zhang, and Y. Zhang, "Multimodal emotion recognition from EEG signals and facial expressions," IEEE Access, vol. 11, pp. 23456-23468, 2023, [Online]. Available: https://doi.org/10.1109/access.2023.3263670.

[18] R. Tshibangu and J. R. Tapamo, "Improving facial emotional recognition using convolution neural network with minimal layers," in Electronics, Communications, and Networks, Amsterdam, The Netherlands: IOS Press, 2024, pp. 672-682, [Online]. Available: https://doi.org/10.3233/FAIA231252.

[19] J. Pan, W. Fang, Z. Zhang, B. Chen, Z. Zhang, and S. Wang, "Multimodal emotion recognition based on facial expressions, speech, and EEG," IEEE Open J. Eng. Med. Biol., vol. 5, pp. 396-403, 2024, [Online]. Available: https://doi.org/10.1109/OJEMB.2023.3240280.

[20] Z. Wu and D. Pan, "The application and optimization of deep learning in recognizing student learning emotions," Trait. Signal, vol. 41, no. 1, pp. 391-399, 2024, [Online]. Available: https://doi.org/10.18280/ts.410133.

[21] Face expression recognition dataset, Kaggle, 2017, [Online]. Available: https://www.kaggle.com/datasets/jonathanoheix/face-expression-recognition-dataset.

[22] Facial emotion recognition dataset, Kaggle, 2018, [Online]. Available: https://www.kaggle.com/datasets/chiragsoni/ferdata

[23] Facial Emotion dataset, Kaggle, 2019, [Online]. Available: https://www.kaggle.com/datasets/mh0386/facial-emotion.