SemantJournals American Journal of Technology Advancement 2997-9382, Vol 03, No. 04, 2026

Article

An Attention-Enhanced Deep Learning Algorithm for Emotion Pattern Mining in Facial Recognition

Nadia Mahmood Ali
Institute of Medical Technology, Al-Mansur, Middle Technical University, Baghdad, Iraq
nadiam@mtu.edu.iq

Munera A. Jabaar
Middle Technical University - Al-Rusafa Management Institute, Iraq
munera_abdaljabar@mtu.edu.iq

Abstract: Facial emotion recognition (FER) poses continuous algorithmic difficulty because of facial expression variations, occlusion, and lighting as well as class imbalance. The current paper presents AMACNN, an attention-enhanced deep learning (DL) algorithm, which is specific to emotion pattern mining using static facial images. The suggested model integrates multi-scale convolutional streams and spatial attention for extracting local and global features and target emotion-relevant face regions. In order to counter the problem of data imbalance, we use a hybrid loss function that is a combination of weighted cross-entropy loss and focal loss to make the model more sensitive to under-represented classes. Assessed on three big-scale databases of facial emotions (more than 100,000 images) the AMACNN reaches its highest accuracy of 95.4% and its precision, recall, and F1-scores are balanced across all the emotions classes. These findings support the generalizability and robustness of the algorithm and makes it one of the competitive FER solutions to be used in computer vision and affective computing in real-world applications.

Keywords: Deep Learning, Facial Emotion Recognition, Computer Vision, Emotion Analysis, Machine Learning, Data Mining, Adaptive algorithms, Image processing, Artificial intelligence.

1. Introduction

FER has become a key area of intersection between computer vision and affective computing as a result of the increasing number of applications in need of an intelligent system that can interpret human emotions through visual cues. The systems have been permeating into application in human-computer interaction (HCI), medical diagnostics, surveillance, and customized services [1][2][3][4]. The emergence of DL and hybrid optimization techniques has made the interpretation of visual data very advanced across fields, healthcare included with secure and efficient medical image retrieval systems constructed on convolutional networks and Pareto type optimization [5]. These trends highlight the growing need to have adaptive, domain-sensitive algorithms in real-world image analysis- facial emotion recognition. Recent developments in DL, especially Convolutional Neural Net (CNNs), have greatly enhanced the performance of FER by allowing automatic extraction of complex hierarchical features of facial images. In comparison to the traditional techniques where the features are handcrafted, DL methods are more robust to changes in facial expressions, head position, light, and occlusions [6]. Using the pre-defined features and statistical models in processing facial expressions created suboptimal outcomes because of the large range of natural facial expressions systems such as variations in lighting, partial reveals of faces, head position decisions, and differences in cultural presentations. The direct learning capability of DL models allows them to extract complex, robust image representations out of unprocessed information and addresses many of the traditional constraints of analysis [7][8][9][10]. Nevertheless, there are still a number of challenges even in the midst of these developments. Models should correspond to the various levels of accuracy and be able to operate in various conditions of the real world to make the variability of facial expressions as a result of individual differences and environmental factors [11,12]. The absence of significant and labeled datasets in specific domains leads to overfitting of the model, undermining its capacity to generalize in various circumstances. According to the study conducted by [13] and [14], new model architecture innovations to solve these challenges involve combining attention mechanisms and data augmentation methods to address these issues. The methods serve to highlight the important features on the face besides rendering models insensitive to varying types of data [15,16]. FER has led to tremendous growth in application in various industries as a result of methodological developments that have increased the field of use. Such technologies have recently become fundamental to the operations of business in the healthcare systems, retail businesses, security management as well as entertainment industries [17,18]. Real-time emotional analysis functions offer retail operators and healthcare providers tools to adjust their services to the mood of the clients and ensure patient health conditions. Emotion recognition systems are an additional behavioral analysis tool applied in security applications and functions to identify security threats and quantify the possible threat. Combining the visual inputs with the audio or textual cues prove to be capable of promising recognition of the human emotions, leading to increased access to these systems in more difficult operation areas [19,20].

The major weakness of the existing FER models is their reduced generalizability across different datasets and in the real world. Most of the techniques in use find it challenging:

1. Class imbalance, in which certain emotion categories are under-represented.
2. Sensitivity to the environmental variations, like occlusion, lighting, and facial orientation.
3. Limited capability to focus on most informative facial regions, which decreases recognition accuracy.

These weaknesses result in poor performance when it comes to real life applications, especially in uncontrolled environments. To overcome them, this paper proposes a new DL model the Adaptive Multi-scale Attention-based Convolutional Neural Network (AMACNN). The suggested framework will focus on:

1. Integrating multi-scale feature extraction for capturing both fine-grained as well as global facial details.
2. Applying spatial attention mechanism for dynamically focusing on the face's critical regions (mouth, eyes, eyebrows).
3. Using hybrid loss function combining weighted cross-entropy as well as focal loss for mitigating class imbalance, also enhancing performance on under-represented emotions.

The major aims of the presented work:

1. Designing and implementing a robust DL model (AMACNN) with regard to static FER.
2. Evaluating the efficiency regarding multi-scale convolution as well as spatial attention in enhancing the accuracy of FER.
3. Assessing the generalization of the model across multiple facial expression datasets.
4. Comparing the performance of AMACNN with state-of-the-art FER models utilizing standard evaluation metrics.

The structure of this paper is based on the related work analysis, the description of a proposed method, the results and their analysis, as well as the final section concerning the conclusions and research directions. Figure 1 presents a list of the typical facial expressions.

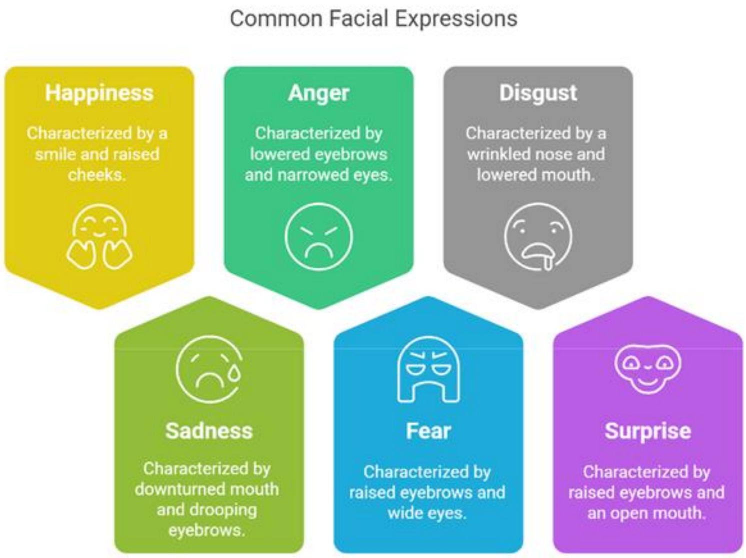


Figure 1. List Of Popular Facial Expressions

Related Work

Various studies have dealt with the FER with the use of DL. The original approach has been developed by Ngo and Yoon [21] using the concept of weighted-cluster loss and deep transfer learning. They demonstrated that their method achieved competitive accuracy even though the dataset used was highly imbalanced. The imbalance in dataset can however have an impact on the model generalizability. Oh et al. [22] designed a driver emotion recognizer, which is a DL model. In their approach, they were able to capture facial expression to gauge the emotions of the driver, which showed a possibility of road safety benefits. Nevertheless, the strength of the model to different lighting conditions and the environment was not completely discussed. Alonazi et al. [23] suggested an automated FER system based on Pelican Optimization Algorithm coupled with CNN, with a competitive accuracy. It is, however, limited by its metaheuristic tuning that could decrease the ability to adapt to unseen data. Punuri et al. [24] created EfficientNet-XGBoost on FER with a transfer learning, and it performs better, although its generalization beyond the pretrained domains is problematic. Cîrneanu et al. [25] provided a systematic review of emotion recognition tendencies on neural networks (NNs) where more robust adaptive models should be used in unconstrained conditions. Mukhiddinov et al. [26] concentrated on masked-face emotion recognition on the basis of facial landmarks and CNNs to assist visually impaired individuals, whereas the general FER was not emphasized on a variety of datasets. Chowdhury et al. [27] indicated that VGG model accuracy in FER could be enhanced by simple histogram equalization, yet the architectural innovation has been limited. Xuan and Vinh [28] investigated the concept of facial sentiment recognition in 2023, entailing the combination of the artificial intelligence (AI) methods with hardware systems. Their solution is new, but it has not been empirically tested in practice widely. Zhang [29] enhanced VGG-16 model to detect facial emotions by means of the FER dataset. This model demonstrated a good recognition ability. However, relying on single dataset could restrict its effectiveness in diverse situations. Chen and Zhang [30] presented a new algorithm of FER using deep neural networks (DNNs) which was developed by a new loss function to address the issue of data imbalance. Even though the approach has potential, its complexity may not support real-time implementation. Wang et al. [31] were interested in the multimodal recognition of emotions, integrating EEG signal with facial expression. Their DL model obtained the excellent results in classification by means of efficient multimodal integration. Nonetheless, the use of EEG data can reduce the applicability of the model in situations where the EEG data are not provided. Sadahide et al. [32] carried out a clinical trial to test the DL-facilitated facial recognition as an effective method of patient identification in hospitals. Their approach was very accurate in unmasked scenarios. The research however did not deal with the issues of masked faces. Chouhayebi et al. [33] suggested a technique that operates on spatiotemporal face features that are extracted using HOG-HOF and VGG-LSTM. This makes their method more robust in feature extraction. Its main weakness is that it is computationally intensive and this might act as a hindrance to real-time application. ConvNet was created by Tshibangu and Tapamo [34], and it is a classification model that can extract features of a facial image using different methods. This approach improved the level of recognition; however, its complex structure is a practical implementation issue. The study conducted by Pan et al. [35] outlines the Deep-Emotion model which combines facial expressions with speech and EEG data to make a multimodal fear recognition system. Combination of the discriminative features is adaptive to their method of enhancing recognition performance. This model can be deployed in real-time where it presents a challenge because of its technical complexity. Wu and Pan [36] attempted to improve the methods of DL to identify the emotional states of student learning. The study has effectively addressed the issues of varying lighting conditions and particular participant factors. The model is limited in that it is specific to a given population. A short comparison of the related works under discussion regarding the strengths and weaknesses of the corresponding works is given in Table 1.

Table 1. Summary Of Related Works Discussed

Ref.	Approach (Method Used)	Strength Points	Weak Points or Restrictions
[21]	Weighted-cluster loss with deep transfer learning	Achieved competitive accuracy despite dataset imbalance	The imbalanced dataset may affect the generalizability
[22]	Deep learning–based driver's emotion recognizer	Effectively captures driver emotions; potential road safety benefits	Limited robustness under varying lighting conditions
[23]	An automated FER system using the Pelican Optimization Algorithm integrated with CNN	Achieving competitive accuracy	Reliance on metaheuristic tuning may limit adaptability to unseen data
[24]	EfficientNet-XGBoost for FER using transfer learning	Improved performance over conventional methods	Generalization outside pretrained domains remains a challenge.
[25]	Systematic review on emotion recognition trends using neural networks	Significant accuracy improvements	Lacks detailed computational cost analysis
[26]	Masked-face emotion recognition using facial landmarks and CNNs to support visually impaired individuals	Efficient extraction of facial features	Lacked emphasis on general FER in diverse datasets.
[27]	Simple histogram equalization	Enhance VGG model accuracy	Architectural innovation was limited
[28]	AI-based facial sentiment recognition with hardware integration	Innovative integration of AI and hardware	Lacks extensive real-world validation
[29]	Enhanced VGG-16 model using FER-2013 dataset	Effective emotion recognition	Dependency on a single dataset
[30]	Deep learning algorithm with novel loss function for data imbalance	Promising in addressing data imbalance	Complexity may hinder real-time application
[31]	Multimodal emotion recognition integrating EEG and facial expressions	Excellent classification through multimodal integration	Reliance on EEG data limits broader applicability

Ref.	Approach (Method Used)	Strength Points	Weak Points or Restrictions
[32]	Deep learning-based facial recognition for patient identification in clinical settings	High accuracy in clinical trials	It does not address masked face recognition
[33]	Spatio-temporal feature extraction using HOG-HOF and VGG-LSTM	Robust feature extraction	High computational demands
[34]	Classification model that extracts features from facial images using multiple techniques	Improved FER performance through multiple techniques	Complex architecture may restrict practical use
[35]	Multimodal system integrating facial expressions, speech, and EEG (Deep-Emotion model)	Adaptive fusion of features enhances recognition performance	Model complexity may hinder real-time deployment
[36]	Optimized deep learning for student learning emotion recognition	Effectively address lighting and individual differences	Specific demographic focus reduces generalizability

2. Materials and Methods

The present study employs a numerical-analytical method to postulate the thermal performance of heat exchangers by nanofluids under turbulent flow. Methodology designed to simulate realistic operating conditions and ensure that thermal and hydraulic behaviors are accurately recorded. The nanofluid is a single-phase homogeneous liquid, consisting of the base liquid and nanoparticles which are in thermal equilibrium with each other. According to this assumption is widely used in engineering applications due to its simplicity and holds reasonable accuracy at low to moderate concentrations of nanoparticles. [9]

In this paper, a new DL framework, namely, the AMACNN is proposed as a FER model and some new aspects are brought in to enhance the recognition accuracy and robustness. Figure 2. flowchart of the proposed AMACNN algorithm. The elements of the proposed model are presented in Figure 3.

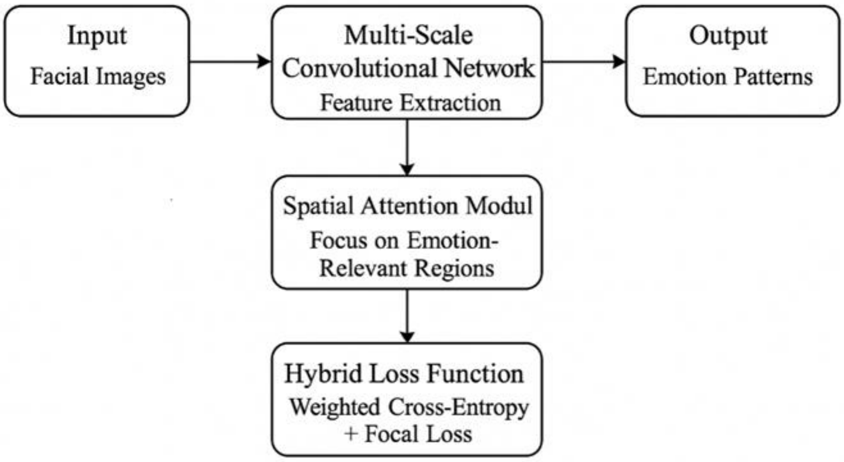


Figure 2. Flowchart of The Proposed Model

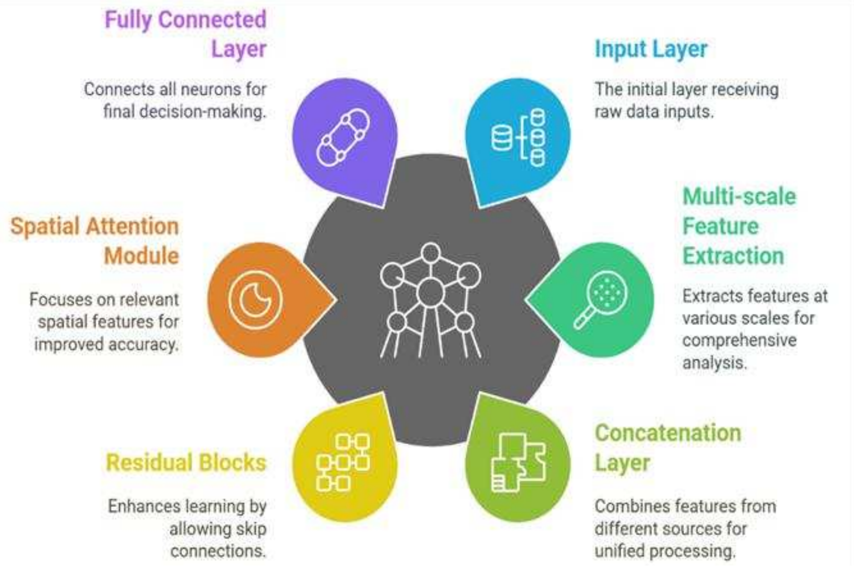


Figure 3. The Components of the Proposed Model

The AMACNN structure consists of the following components (see Figure 3):

1. Multi-scale convolutional streams for extracting features at different spatial resolutions.
2. Residual refinement blocks to improve feature representation and ensure gradient stability.
3. A spatial attention module emphasizing emotion-relevant facial regions.
4. Fully connected layers as well as SoftMax classifier for final emotion categorization.

Figure 4 illustrates the proposed methodology.

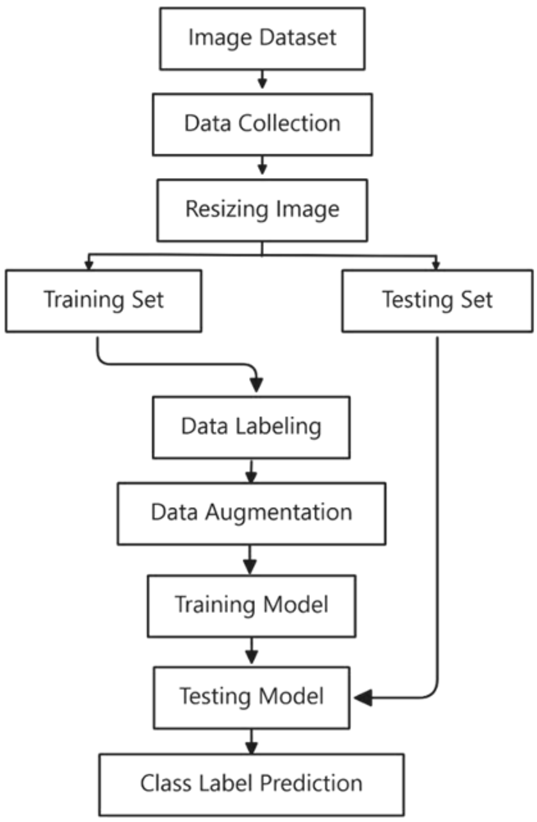


Figure 4. Schematic Diagram of the Proposed Methodology

Data collection involves a uniform image resizing of 128×128 pixels, as well as retaining key facial features with minimal input data processing computational demands. An 80:20 split technique is used to divide the data into the training and testing parts to have sufficient data to build a solid model and ensure objective evaluation of the model results. Data augmentation is directed at the training set alone, where geometric and photometric modifications alongside simulations of occlusions are made in order to raise the diversity of datasets and avoid overtraining. The hybrid loss function that combines weighted cross-entropy with focal loss helps manage class imbalance by prioritizing minority classes while adjusting the weighted loss according to well-predicted outputs. The training process optimizes models through Adam optimization, dropout techniques, L2 regularization, and early stopping to stop overfitting. Figure 5 provides the architectural design of the proposed model.

Multi-scale Feature Extraction

A multi-scale feature extraction module in the proposed method extracts different facial expression patterns across different image resolution scales. The network converts input images through parallel convolutional streams operating with kernel sizes of 3×3, 5×5 and 7×7. Its network architecture offers the capability to retrieve small-scale muscular expressions and overall facial information with contextual value. The multiple streams receive added nonlinear activation features with batch normalization to reach stable learning and faster convergence. Features produced from parallel streams merge into a sequence of residual blocks based on ResNet designs for additional feature refinement while maintaining gradient flow. The model requires multiple scales of information processing to deal with illumination variation, occlusions, and different facial poses since such issues proved challenging in previous research approaches.

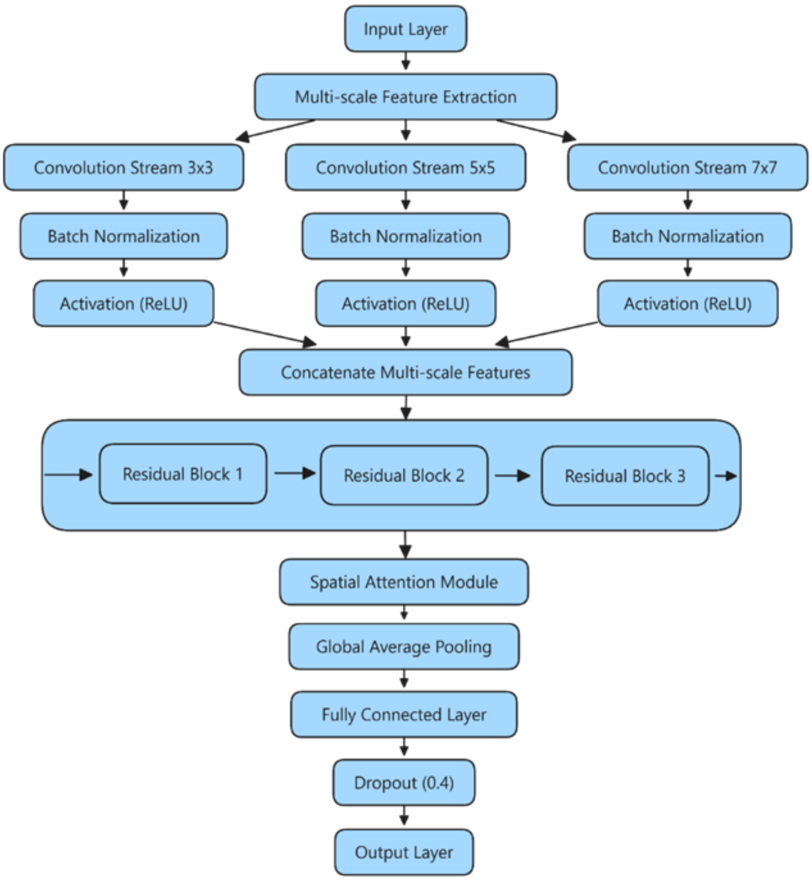


Figure 5. The Architecture of the Proposed Model

Spatial Attention Mechanism

The AMACNN is further used to discriminate features by adding spatial attention mechanism. This element focuses on parts of the face that are most informative in emotions recognition, including the mouth, eyes, and eyebrows, and inhibits the rest of the background information that is irrelevant.

The attention module is used as an assistant network branch which calculates the weights of attention on the concatenated multi-scale feature maps. These weights are then applied to the feature maps element-wise so that the network can concentrate its representational power on essential parts of the face. Incorporating the aspect of attention enhances the performance of the classification and forms part of the model strength as it mitigates the influences of extraneous differences in the input images. It is especially relevant when working with datasets which can have images of different environmental conditions.

The complete algorithm of the suggested AMACNN scheme may be outlined by Algorithm 1:
Algorithm1: General steps for the neural network architecture

Algorithm1: General steps for the neural network architecture
Input: Pre-Processed image Output: Classified image 1. Data Acquisition and Preprocessing a. Acquire a diverse, multi-domain dataset of static facial images. b. Normalize image dimensions and apply standard preprocessing (mean subtraction, scaling). c. Execute data augmentation techniques to generate a robust training set. 2. Multi-scale Feature Extraction a. Input images are fed into parallel convolutional streams with varying kernel sizes. b. Apply batch normalization and ReLU activations in each stream. c. Concatenate the outputs from all streams. 3. Residual Feature Refinement a. Pass concatenated feature maps through a series of residual blocks to refine and consolidate features. b. Maintain gradient flow and stabilize training through residual connections. 4. Spatial Attention Integration a. Compute attention weights over the refined feature maps using an auxiliary attention network. b. Apply the computed attention weights to the feature maps to emphasize critical regions. 5. Feature Fusion a. Fuse the attention-enhanced feature maps and flatten the output. b. Process the fused features through fully connected layers with dropout for regularization. 6. Classification a. Use a SoftMax layer to classify the emotion into one of the predefined categories.

The hyperparameter values in the experiments are summed up in Table 2. A summary of the network architecture parameters is given in table 3.

Table 2. Hyperparameter Values Used

Parameter	Value	Description
Learning Rate	0.0005	The initial learning rate for the Adam optimizer
Batch Size	32	Number of samples per training batch
Epochs	45	Total number of training iterations

Parameter	Value	Description
Dropout Rate	0.4	Dropout probability for fully connected layers
Weight Decay	0.0001	L2 regularization coefficient to reduce overfitting
Focal Loss Gamma	2.0	Focusing parameter for the focal loss component
Focal Loss Alpha	0.25	Balancing parameter for the focal loss component

Table 3. Summary Of Network Structural Parameters

Layer Type	Parameters	Description
Input Layer	128×128×3	Standardized input image dimensions
Convolutional Layer (Stream 1)	64 filters, 3×3 kernel, stride 1	Extracts fine-grained features from the input image
Convolutional Layer (Stream 2)	64 filters, 5×5 kernel, stride 1	Extracts mid-level features; larger receptive field
Convolutional Layer (Stream 3)	64 filters, 7×7 kernel, stride 1	Extracts global contextual features
Concatenation Layer	N/A	Merges feature maps from the parallel streams
Residual Blocks	3 blocks; each with 2× (Conv + ReLU)	Refines and consolidates multi-scale features
Attention Module	1×1 convolutions for attention weighting	Generates spatial attention maps to emphasize key facial regions
Fully Connected Layer	256 units	Integrates high-level features before classification
SoftMax Output Layer	Number of emotion classes	Produces probability distribution over emotion categories

The AMACNN system has a number of novel contributions to FER. Its multi-scale analysis takes advantage of parallel convolutional streams that have different receptive fields to make the model capture both the local and global features. Also, by incorporating an attention-based feature refinement module, the spatial attention mechanism can dynamically prioritize important facial areas and, therefore, increase the discriminatory strength of the extracted features. The framework uses a hybrid loss, which is a weighted combination of cross-entropy and focal loss, which is useful in controlling the imbalance of classes and enhancing recognition performance on emotion classes that are underrepresented. Robust data augmentation methods are also provided, which makes the model resistant to the changes in light and occlusion, as well as pose typical problems in the real world. Lastly, the combination of the Adam optimizer, dropout and L2 regularization offer effective training and reduces the threat of overfitting. The AMACNN structure will be a major breakthrough in stat FER, as it will serve as a contribution to earlier methods and combine the new elements of the process.

3. Results and Discussion

Experiments on the suggested AMACNN framework were carried out on a powerful workstation that has an Intel i9 processor and a 32 GB of RAM and NVIDIA RTX 3080 graphics card. The implementation environment was based on the use of version 2.9 of TensorFlow with Keras API and OpenCV to process and process images, respectively, to compute performance metrics using sci-kit-learn. Training was allocated 80% and 20% was allocated to testing. The model was trained on three various datasets available on Kaggle: The first Facial Expression Dataset has approximately 36,000 images, the Facial Recognition Dataset has approximately 35,000 images, and the second Facial Expression Dataset has approximately 30,000 images. This training extension introduced the strength of robustness by applying data augmentation methods, which comprised different geometric transformations as well as photometric adjustments and simulated occlusions. To enhance the trade-off between the weighted cross-classes and focal loss a hybrid loss was applied. The results of evaluation indicate that AMACNN system has superior performance to the traditional techniques. The model has been effective to the extent that it accommodates attention mechanisms and multi scale features to identify local and general facial features effectively resulting in the effectual generation of results with varying datasets. The normative evaluation outcomes provided in Table 4 and Figure 6 support the idea that the model had excellent performance measures, in three face expression datasets, including precision, recall, and F1 scores, which are more or less similar. The model shows good generalization ability in its performance data when it is operating in changes of lighting pose and occlusion conditions.

Table 4. Performance Metrics on Different Datasets

Dataset	Ref.	Accuracy	Precision	Recall	F1-Score
1st Facial Expression Dataset (36K images)	[37]	95.4%	95.3%	95.5%	95.4%
Facial Recognition Dataset (35K images)	[38]	94.7%	94.6%	94.8%	94.7%
2nd Facial Expression Dataset (30K images)	[39]	94.8%	94.8%	94.9%	94.8%

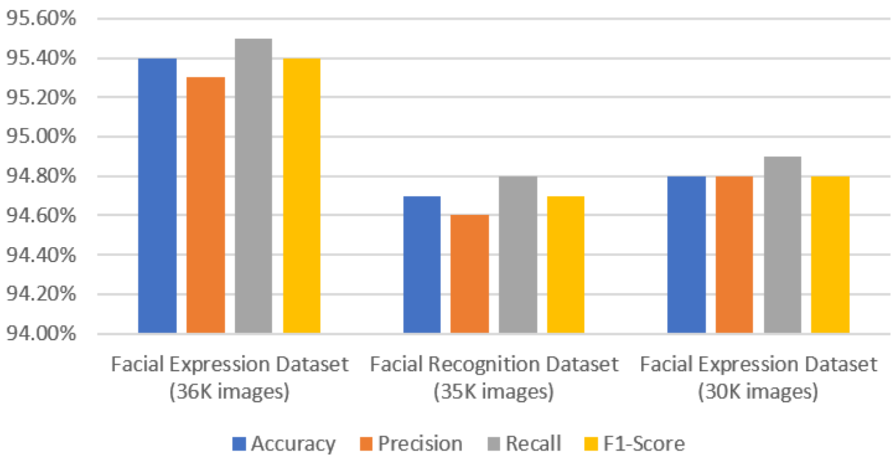


Figure 6. Performance Metrics on Three Datasets

The related techniques that were critically compared have shown the superiority of the offered framework over them. The AMACNN is the best in that it performed well in its precision of results, as well as balanced metrics performance during all assessments. Combining hybrid loss function, parallel convolutional streams, a practical attention module, the state-of-the-art data augmentation enhance the model performance. The training progress of the accuracy levels and loss measurements is illustrated in Figure 7.

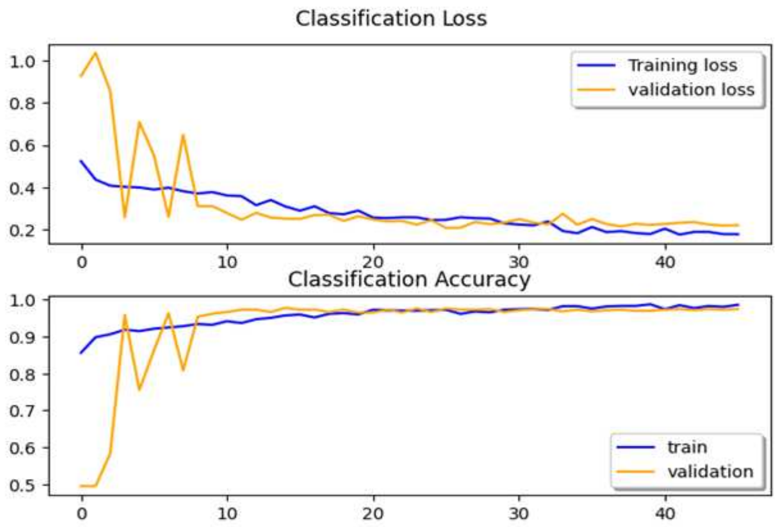
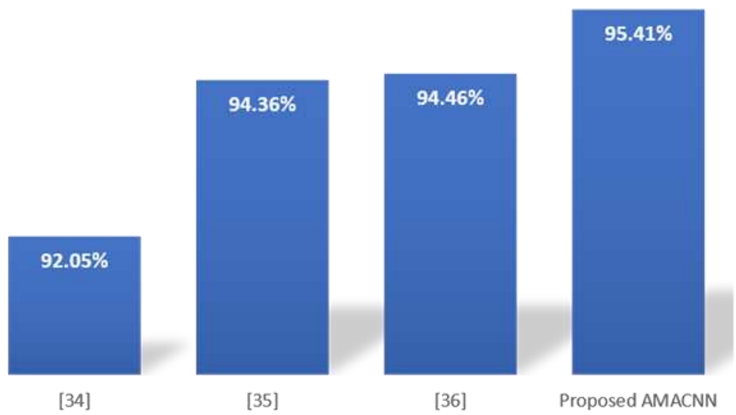


Figure 7. The Training Curves for Accuracy and Loss

Table 5 and Figure 8 show the evaluation of proposed AMACNN and traditional methods. The reason why AMACNN framework delivers excellent outcomes is due to the fact that it is better than conventional approaches and yields higher accuracy rates and precision indicators, as well as recall and F1-score. The approach has better functionalities that can prove its efficiency in successfully handling the emotional facial recognition issues.

Table 5. Comparison With Existing Methods

Method	Accuracy
[34]	92.05%
[35]	94.36%
[36]	94.46%
Proposed AMACNN	95.41%



The analysis and the results of the experiment confirm that the AMACNN framework provides the top-level performance in recognizing facial emotions. The innovative architecture elements and robust training strategy would be an effective approach to the issues of data imbalance, overfitting, and computational complexity. This comprehensive method makes the AMACNN a prospective candidate of practical applications and creates a new standard of research in the future.

4. Conclusion

This paper introduced AMACNN which is a DL framework for static FER combining multi-scale feature extraction with dynamic spatial attention mechanism. We use multi-scale feature extraction and dynamic spatial attention to identify fine-grained features and global elements to enhance the discriminative ability of the model. The issues of class imbalance were addressed by the combination of weighted cross-entropy and focal loss as a hybrid loss function to deliver balanced performance between emotions.

The Facial Expression Dataset of some 36,000 images showed that the experimental model achieved an overall accuracy of 95.4% and consistent results between precision and recall and F1-scores. The framework is proposed to operate in a wide range of environmental conditions, such as the different lighting environments, facial poses, and partial disruptions. The model has excellent generalizability since it incorporates data augmentation in conjunction with the use of advanced optimization features, such as the Adam Optimizer, dropout, and L2 regularization, preventing the model to overfit. In comparative testing, AMACNN is better than the conventional practices and establishes a new standard in the field.

Additional research opportunities exist to integrate more modalities into the framework and apply the method to real-time applications after its recent improvements. The AMACNN framework marks a substantial advance in efficient facial emotion recognition that shows promising utility for practical applications in affective computing and human-computer interaction.

Future research may explore:

1. Real-time deployment using lightweight variants of AMACNN,
2. Integration of multimodal inputs (e.g., speech, body language),
3. Extension to dynamic (video-based) facial emotion recognition.

This work sets a strong foundation for the development of reliable and scalable FER systems for applications in affective computing, healthcare, education, and security.

Author Contributions

N.M.A.: Conceptualization, Methodology, Investigation, Writing—original draft.

S.A.L.: Conceptualization, Investigation, Project administration, Conceptualization.

D.H.H.:Software, Investigation,

A.G.H.R.: Conceptualization, Investigation, Writing—review and editing, Resources.

All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

Ethics Statement

This study utilized publicly available datasets—Facial Expression Dataset by Aaditya Singhal, Face Recognition Dataset by Vasu Patel, and FER-2013 Dataset—which are widely used for academic research in facial expression and emotion recognition. No personal, sensitive, or identifiable information was collected, and all data was fully anonymized. The research did not involve any direct interaction with human subjects or animals. Accordingly, institutional ethical approval was not required. This study adheres to the ethical standards outlined for research using open-access data.

Funding

This research received no external funding.

REFERENCES

[1] H. Ge, Z. Zhu, Y. Dai, B. Wang, and X. Wu, “Facial expression recognition based on deep learning,” Comput. Methods Programs Biomed., vol. 215, p. 106621, 2022.

[2] 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, 2023.

[3] M. A. H. Akhand, S. Roy, N. Siddique, M. A. S. Kamal, and T. Shimamura, “Facial emotion recognition

using transfer learning in the deep CNN," Electronics, vol. 10, no. 9, p. 1036, 2021.
[4] S. Min et al., "Emotion recognition using transformers with masked learning," arXiv preprint arXiv:2403.13731, 2024.
[5] S. A. Lafta et al., "Secure medical image retrieval using fast image processing algorithms," Scalable Comput.: Pract. Exp., vol. 25, no. 5, 2024, doi: 10.12694/scpe.v25i5.3126.
[6] P. Karamaouna et al., "Facial emotion recognition and its associations with psychological well-being across four schizotypal dimensions," Arch. Clin. Neuropsychol., 2025.
[7] B. Devi and M. M. S. J. Preetha, "Facial emotion recognition using convolutional neural network-based krill head optimisation," Expert Syst., vol. 42, no. 1, e13376, 2025.
[8] H. Chi, "Facial sentiment recognition using artificial intelligence techniques," EAI Endorsed Trans. Context-Aware Syst., 2023.
[9] M. C. Gursesli et al., "Facial emotion recognition through custom lightweight CNN model," IEEE Access, 2024.
[10] Y. Fu et al., "Blindfold attention: Novel mask strategy for facial expression recognition," Proc. ACM Comput. Graph. Interact. Tech., 2022.
[11] N. Mehendale, "Facial emotion recognition using convolutional neural networks (FERC)," SN Appl. Sci., vol. 2, no. 3, p. 446, 2020.
[12] X. Wang et al., "Complex emotion recognition via facial expressions," Comput. Intell. Neurosci., 2023.
[13] M. Aly, A. Ghallab, and I. S. Fathi, "Enhancing facial expression recognition system in online learning," IEEE Access, vol. 11, pp. 121419–121433, 2023.
[14] I. Lasri, A. Riadsolh, and M. Elbelkacemi, "Facial emotion recognition of deaf students," Educ. Inf. Technol., vol. 28, no. 4, pp. 4069–4092, 2023.
[15] K. Sarvakar et al., "Facial emotion recognition using convolutional neural networks," Mater. Today Proc., vol. 80, pp. 3560–3564, 2023.
[16] Y. Zhang and T. Tan, "Deep emotion recognition in dynamic data," TechRxiv, 2021.
[17] M. M. Husein and D. Alzubaydi, "Mobile face recognition using eigenface approaches," Al-Mustansiriyah J. Sci., vol. 30, no. 1, pp. 1–12, 2019.
[18] M. Rouast et al., "Deep learning for human affect recognition," IEEE Trans. Affect. Comput., 2021.
[19] H. Xia, L. Lu, and S. Song, "Feature fusion for facial expression recognition," Vis. Comput., vol. 40, no. 3, pp. 2035–2047, 2024.
[20] K. Jagadeesh and G. Baranidharan, "Dynamic FERNet for face expression recognition," Concurrency Comput., 2022.
[21] Q. Ngo and S. Yoon, "Facial expression recognition using weighted-cluster loss," Sensors, vol. 20, no. 9, p. 2639, 2020, doi: 10.3390/s20092639.
[22] G. Oh et al., "Driver's real emotion recogniser," Sensors, vol. 21, no. 6, p. 2166, 2021, doi: 10.3390/s21062166.
[23] M. Alonazi et al., "Automated facial emotion recognition using pelican optimization," Electronics, vol. 12, no. 22, p. 4608, 2023, doi: 10.3390/electronics12224608.
[24] S. B. Punuri et al., "EfficientNet-XGBoost for facial emotion recognition," Mathematics, vol. 11, no. 3, p. 776, 2023, doi: 10.3390/math11030776.
[25] A.-L. Cîrneanu, D. Popescu, and D. Iordache, "New trends in emotion recognition," Sensors, vol. 23, no. 16, p. 7092, 2023, doi: 10.3390/s23167092.
[26] M. Mukhiddinov et al., "Masked face emotion recognition using deep learning," Sensors, vol. 23, no. 3, p. 1080, 2023, doi: 10.3390/s23031080.
[27] J. H. Chowdhury, Q. Liu, and S. Ramanna, "Histogram equalization improves FER performance," Algorithms, vol. 17, no. 6, p. 238, 2024, doi: 10.3390/a17060238.
[28] V. Xuan and P. Vinh, "Facial sentiment recognition using AI techniques," EAI Endorsed Trans., vol. 9, no. 1, 2023, doi: 10.4108/eetcasa.v9i1.3930.
[29] R. Zhang, "Facial emotion detection based on improved VGG-16," Appl. Comput. Eng., vol. 21, no. 1, pp. 86–92, 2023, doi: 10.54254/2755-2721/21/20231110.
[30] Y. Chen and M. Zhang, "Face emotion recognition algorithm based on deep learning," Appl. Math.

Nonlinear Sci., vol. 9, no. 1, 2023, doi: 10.2478/amns.2023.2.00533.
[31] S. Wang et al., “Multimodal emotion recognition from EEG and facial expressions,” IEEE Access, vol. 11, pp. 33061–33068, 2023, doi: 10.1109/access.2023.3263670.
[32] A. Sadahide et al., “Deep learning-based facial recognition for patient identification,” Bioengineering, vol. 11, no. 4, p. 384, 2024, doi: 10.3390/bioengineering11040384.
[33] H. Chouhayebi et al., “Human emotion recognition using spatio-temporal features,” Computers, vol. 13, no. 4, p. 101, 2024, doi: 10.3390/computers13040101.
[34] R. Tshibangu and J. R. Tapamo, “Improving facial emotion recognition using CNN,” in Electronics, Communications and Networks, IOS Press, 2024, pp. 672–682.
[35] J. Pan et al., “Multimodal emotion recognition using facial, speech, and EEG,” IEEE Open J. Eng. Med. Biol., vol. 5, pp. 396–403, 2024, doi: 10.1109/ojemb.2023.3240280.
[36] Z. Wu and D. Pan, “Deep learning in recognizing student learning emotions,” Traitement du Signal, vol. 41, no. 1, pp. 391–399, 2024, doi: 10.18280/ts.410133.
[37] A. Singhal, “Facial expression dataset,” Kaggle, 2020. [Online]. Available: https://www.kaggle.com/datasets/aadityasinghal/facial-expression-dataset
[38] V. Patel, “Face recognition dataset,” Kaggle, 2020. [Online]. Available: https://www.kaggle.com/datasets/vasukipatel/face-recognition-dataset
[39] I. Goodfellow et al., “Challenges in representation learning: Facial expression recognition challenge,” Kaggle, 2013. [Online]. Available: https://www.kaggle.com/datasets/msambare/fer2013