2024 IEEE 21st International Conference on Smart Communities: Improving Quality of Life using AI, Robotics and IoT (HONET) | 979-8-3503-7807-8/24/$31.00 ©2024 IEEE | DOI: 10.1109/HONET63146.2024.10822916

Enhancing Micro-Expression Recognition with Broadbent Attention Mechanism: A High-Performance Approach for Emotion Detection

1st Imran Ashraf
Faculty of Computer Science
University of Lahore
Lahore, Pakistan
imranashraf.yaseen@gmail.com

2nd Tanzila Kehkashan
Faculty of Computing
Universiti Teknologi Malaysia
Johor Bahru, Malaysia
tanzila.kehkashan@gmail.com

3rd Muhammad Zaman
Department of Computer Science
COMSATS University Islamabad
Islamabad, Pakistan
mzamancui@gmail.com

4th Adnan Akhunzada
Department of Data and Cybersecurity
University of Doha for Science and Technology
Doha, 24449, Qatar
adnan.akhunzada@udst.edu.qa

5th Hafiz Khizer bin Talib
School of Mechanical Engineering
Zhejiang University
Hangzhou, China
12125129@zju.edu.cn

6th Yanlong cao
School of Mechanical Engineering
Zhejiang University
Hangzhou, China
sdcaoyl@zju.edu.cn

***Abstract*—Micro-expression recognition (MER) is a critical task in affective computing, yet it remains challenging due to the subtle, brief, and involuntary nature of micro-expressions. Existing methods often find it difficult to spot these quick micro facial expressions (MFEs), especially when there are obstacles, background noise, or differences between people. In this study, we propose a novel approach that integrates the Broadbent Attention Mechanism into a Convolutional Neural Network (CNN) to enhance MER. This attention mechanism is designed to focus on the most informative regions of facial images, improving the model's ability to extract relevant features and discard irrelevant information. The model is evaluated on the SAMM dataset, covering various emotions. Extensive experiments are conducted using both K-fold and Stratified K-fold cross-validation techniques to ensure a robust evaluation process. The proposed method achieves high accuracy 99.72%, precision 99.73%, Recall 99.72% and f1-Score 99.72% with Stratified K-fold & accuracy 99.65%, precision 99.65%, Recall 99.66% and f1-Score 99.65% with K-fold, outperforming baseline models and demonstrating its effectiveness in recognizing subtle micro-expressions. These results indicate that the Broadbent Attention Mechanism significantly enhances accuracy and robustness of MER systems, offering a promising solution for applications in human-computer interaction (HCI) and emotion analysis.**

***Index Terms*—Broadbent Attention Mechanism, CNN, Micro-expression Recognition, Stratified K-Fold Cross-validation, SAMM**

I. Introduction

Micro-expression recognition (MER) is a field of study within affective computing and computer vision (CV) that focuses on identifying subtle, involuntary facial expressions that occur within a fraction of a second, typically lasting 1/25 to 1/5 of a second [1]. These fleeting expressions can reveal concealed emotions and are often linked to high-stakes situations or emotional suppression. Detecting these micro-expressions requires advanced image processing and machine learning techniques, such as CNNs) and optical flow methods, due to their brief and subtle nature [2]. Applications of MER span various fields, including psychology, security, human-computer interaction (HCI), and mental health diagnosis [3]. The challenge lies in accurately capturing and classifying these minute movements, which are often invisible to the human eye.

MER specifically involves the identification and analysis of rapid, involuntary facial muscle movements that reveal hidden emotions. These expressions are difficult to detect due to their short duration (typically 40-200 milliseconds) and subtle intensity, making them distinct from regular facial expressions [4]. Early research demonstrated that MEs are universal and can indicate emotions such as fear, happiness, or disgust, even when a person is trying to conceal them [5]. With advancements in technology, CV techniques, such as Local Binary Patterns (LBP) and optical flow, combined with deep learning models like CNNs and Long Short-Term Memory (LSTM) networks, have become essential in automating ME detection. Datasets like CASME and SMIC have been developed to train and evaluate these models [6]. The field is particularly significant in lie detection, mental health assessment, and human-computer interaction(HCI), though challenges remain in generalizing across different subjects and facial dynamics.

Many studies, identified the need for larger and more diverse datasets to avoid biased models, but their focus on accuracy often came at the expense of computational efficiency and model interpretability [7]. Additionally, a study combined optical flow with deep learning for stronger performance, but their reliance on optical flow calculations made the system prone to errors when handling subtle or ambiguous facial movements [8]. A study introduced a multimodal approach that improved recognition accuracy by integrating micro facial expressions (MFE) with head movements and

Authorized licensed use limited to: UNIVERSITY TEKNOLOGI MALAYSIA. Downloaded on January 16,2025 at 04:58:00 UTC from IEEE Xplore. Restrictions apply.

979-8-3503-7807-8/24/$31.00 ©2024 IEEE

physiological signals. However, the approach faced challenges in synchronizing data for real-time applications and required additional sensors, which hindered its scalability [9].

MER proves challenging for affective computing and human-computer interaction given the sensitive and swift nature of MEs, which generally last only a fraction of a second and require little muscular movement. Existing approaches [11], [17] frequently fail to identify and understand these ephemeral facial cues in real time, particularly when faced with occlusions, noise, or differences in facial emotions between persons. CNNs [20], [27], [31] failed over dynamic images due to static filter convolve over random pixels and unable to relate spatio-temporal features. To address these issues, we propose using the Broadbent Filter Model of Attention as a solution, with a focus on improving the detection and classification of relevant microexpression features by filtering out irrelevant or distracting facial information, thereby increasing accuracy and efficiency of MER systems.

Our research study have very significant objectives as listed below:

i To develop an efficient and accurate MER model that balances real-time performance with high accuracy.
ii To handle real-world conditions, such as variations in lighting and diverse facial structures.
iii To improve the interpretability of the model, ensuring that the system can be effectively applied in practical, real-time scenarios.

This research contributes to the field by addressing the need for a practical, real-time MER system. Building on previous studies, this work develops a system that balances accuracy with computational efficiency, making it suitable for real-world applications. The contribution aligns with the identified problem of high computational costs and inefficient real-time processing in existing systems. By optimizing model architecture and improving scalability, this research provides a more viable solution for applications such as security, mental health, and HCI, both locally and globally.

The whole paper is organized as follows: *Literature Review* section provides an overview of related work, focusing on recent advancements and their limitations. *Proposed Methodology* section discusses the proposed methodology, including the model architecture, preprocessing techniques, and cross-validation strategies. *Results and Discussion* section presents the results of the study, evaluating the system's performance using various metrics. Finally, *Conclusion* section offers a conclusion and discusses future work, outlining potential directions for further research.

II. Literature Review

Micro Facial Expression Recognition (MFER) has seen remarkable results with the integration of deep learning techniques, leading to enhanced accuracy and efficiency. Early research primarily focused on traditional machine learning techniques, such as feature-based approaches, which laid the groundwork for FER before the deep learning era [10]. These methods relied hardly on handcrafted features and classical algorithms like Support Vector Machines (SVMs) and k-Nearest Neighbors (k-NN), which, while effective, struggled with scalability and real-time performance. As machine learning evolved, these methods gave way to more sophisticated deep learning models, significantly improving FER performance by automating feature extraction [12].

A. Traditional Machine Learning Techniques

Traditional FER methods focused on feature extraction techniques such as the Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP). Lakshmi and Ponnusamy introduced a hybrid approach that combined these methods with deep stacked autoencoders, which were tested on the CK+ and JAFFE datasets [13]. Their approach demonstrated how robust feature extraction could boost emotion recognition, particularly in datasets with varied expressions [14]. These earlier models laid the foundation for more advanced architectures by addressing the challenges of feature selection and recognition in varied lighting and facial structure conditions. However, they often lacked the computational efficiency needed for real-time applications [15].

B. Deep Learning Techniques

With the advent of deep learning, CNNs emerged as the dominant technique for FER. Sharmeen et al. highlighted how CNNs outperformed traditional methods, particularly in handling large public datasets [16]. These models demonstrated improved accuracy and scalability by automating feature extraction. However, the choice of dataset was found to significantly affect model performance, as different databases yielded varying results [18]. Sepas-Moghaddam and Etemad introduced CapsField, a deep learning approach using light field cameras and capsule networks, specifically designed for face and expression recognition in uncontrolled environments [19]. CapsField's ability to capture spatial hierarchies made it more resilient to variations in pose and lighting conditions, outperforming state-of-the-art models. [21] Further advances in deep learning included approaches like the multi-scale atrous convolutional neural network (MACNN), which utilized atrous convolutions to detect facial expressions across different scales [22]. This method proved robust and capable of real-time ME detection on the CASME and CASME II datasets, demonstrating significant improvements in feature extraction. A study proposed the Feature Refinement (FeatRef) method, which enhanced micro-expression recognition by focusing on expression-specific features [23]. These methods emphasized the need for more specialized deep learning models tailored for the subtleties of MEs, a crucial area for advancing FER research.

C. Temporal and Spatiotemporal Models

Recent research has shifted towards recognizing facial expressions over time, capturing both spatial and temporal dynamics. The integration of VGGNets and Long Short-Term Memory (LSTM) networks allowed for the modeling

Authorized licensed use limited to: UNIVERSITY TEKNOLOGI MALAYSIA. Downloaded on January 16,2025 at 04:58:00 UTC from IEEE Xplore. Restrictions apply.

of temporal sequences, capturing subtle changes in facial expressions over time [24]. This approach, validated on the CASME II dataset, demonstrated significant improvements in detecting micro-expressions, which are often too subtle to be captured by static models [25]. Moreover, a study introduced a siamese 3D convolutional neural network, which preserved spatiotemporal information crucial for recognizing spontaneous micro-expressions [26]. Their method offered superior performance on public datasets and highlighted the potential of leveraging scarce data through innovative network designs [36].

In addition, a study proposed a method combining optical flow with deep learning to strengthen FER performance [28]. Although this approach showed promise in handling subtle facial movements, the reliance on optical flow calculations introduced vulnerabilities when dealing with ambiguous or overlapping expressions. Multimodal approaches [29], integrated additional signals like head movements and physiological data to enhance recognition [30]. While this improved accuracy, it introduced challenges regarding data synchronization and scalability for real-time applications.

D. Challenges in Micro-Expression Recognition

Despite the advancements in FER, several challenges persist [32]. Many deep learning models, such as those by Liu et al., achieve high accuracy but require substantial computational resources and large datasets, limiting their real-world applications [33]. Yan et al. improved recognition performance with spatiotemporal recurrent CNNs (STR-CNNs), yet struggled with real-time efficiency and scalability when applied to large datasets [34]. Additionally, methods like the one proposed by Song and Li highlighted the need for more diverse datasets to prevent model bias. However, their emphasis on accuracy often compromised computational efficiency and model interpretability [35].

The literature underscores the need for a more efficient, scalable, and interpretable micro-expression recognition system capable of handling real-world conditions. Our proposed methodology will address these gaps by integrating spatiotemporal feature extraction with real-time performance optimization, while leveraging multimodal approaches to enhance system robustness in diverse environments.

III. Proposed Methodology

In this study, the goal is to develop a model capable of recognizing micro-expressions from facial images, specifically using the SAMM datasets. The approach involves comparing a baseline method with a proposed enhanced model that integrates a novel attention mechanism.

A. Baseline Method

The baseline methodology, HTNet, is a deep learning model for MER using transformer layers with multi-head self-attention and block aggregation. It extracts facial optical flow with Gunnar Farneback's algorithm, targeting four key facial regions to reduce noise. HTNet's hierarchical architecture applies 3 × 3 max pooling across 16 facial blocks and uses as multi-layer perceptron (MLP) for classification with cross-entropy loss. Experiments on CASME II, CASME III, SMIC, and SAMM demonstrate superior performance in UF1 and UAR.Our proposed method enhances HTNet with a novel self-attention mechanism, improving local and global feature capture.

B. Proposed Model

The Broadbent Attention Layer, stimulate by Broadbent's attention model, enhances feature extraction in convolutional neural networks (CNNs) by focusing on essential spatial details through query, key, and value transformations. By filtering out irrelevant information, it boosts classification performance, especially in tasks like micro-expression recognition that rely on detecting subtle facial movements. Following are the steps of our proposed model:

C. Dataset

The experiment uses the SAMM dataset [37], designed for micro-expression recognition, with frames extracted from video sequences undergoing preprocessing. SAMM captures spontaneous emotions like Anger, Disgust, and Surprise in a controlled environment, recorded at 200 fps, allowing detailed analysis of subtle, involuntary facial movements. The high frame rate ensures that fine-grained emotional shifts are captured, enabling effective micro-expression recognition, even without the need for full video analysis. Dataset summary is shown in Table I.

TABLE I
EMOTION DATASETS OVERVIEW

Dataset	Classes	No. of Classes	No. of Frames
SAMM	Anger, Contempt, Disgust, Fear, Happiness, Other, Sadness, Surprise	8	11,722

D. Preprocessing

The images are resized to 64x64 pixels in order to maintain uniformity within the dataset. Pixel values are normalized between 0 and 1. There are different data augmentation techniques, such as alteration of brightness and contrast, used to improve data variability and thus enhance the model's ability to generalize.

E. Model Architecture

Proposed model architecture combines convolutional layers, attention mechanisms, and dense layers to effectively capture and classify micro-expressions from images.

1) Convolutional Layers: The model begins with convolutional layers (16, 32, and 64 filters, 3x3 kernel) to capture spatial features. Each layer uses ReLU activation and batch normalization for stability, allowing the network to learn complex patterns. The architecture is shown in Figure1.

Authorized licensed use limited to: UNIVERSITY TEKNOLOGI MALAYSIA. Downloaded on January 16,2025 at 04:58:00 UTC from IEEE Xplore. Restrictions apply.

2) Broadbent Attention Mechanism: After the convolutional layers, the Broadbent Attention Layer refines feature maps by focusing on important spatial regions, enhancing the model's ability to detect subtle micro-expressions and boosting performance.

3) Global Average Pooling & Dense Layers: Global average pooling reduces feature maps to 1x1x16, minimizing parameters and overfitting. The reduced maps pass through dense layers with ReLU activation and dropout for regularization. A softmax output layer classifies images into eight emotion categories. The diagram illustrates the architecture of

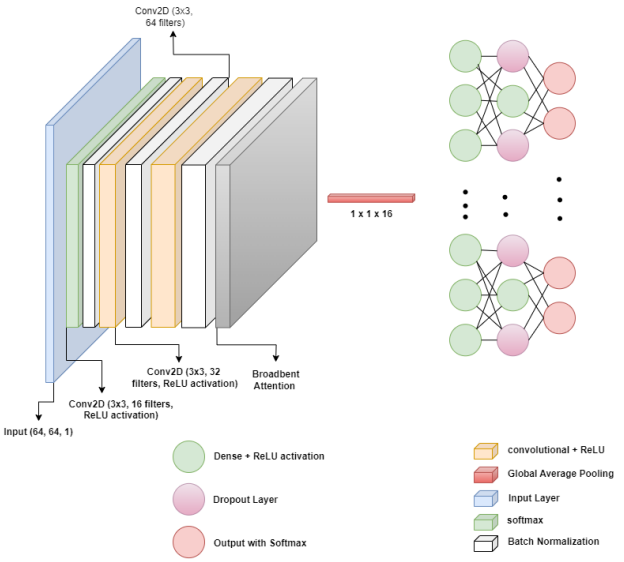

Fig. 1. Proposed Broadbent Attention Mechanism Model Architecture

the proposed Broadbent Attention Mechanism model for MER. The model takes a grayscale image of size 64x64 as input and passes it through a series of convolutional layers with ReLU activation. After convolutional processing (Conv2D), the output is refined using Broadbent Attention, which focuses on important features. The attention-enhanced features are then processed through dense layers with ReLU activations, followed by dropout for regularization, and finally output through a softmax layer for classification.

F. Experiment

In this section, we detail the experimental setup, including the training parameters, dataset specifics, and evaluation methodology used to assess the performance of the proposed model.

1) Training Parameters: Model is trained using a variety of settings, which are fine-tuned through experimentation. The focus is on identifying the best-performing configuration of parameters for each dataset. Table II

IV. RESULTS AND DISCUSSION

This section evaluates the experimental outcomes and assesses model's performance in enhancing MER.

Broadbent Attention Layer method achieved near-perfect

TABLE II
DETAILS OF THE DATASETS AND HYPERPARAMETERS USED IN TRAINING.

Dataset	Cross Validation	No. of Epochs	No. of Folds	Learning Rate	Batch Size	Optimizer Algorithm	SEED	Image Size
SAMM	Stratified K Fold	25	8	0.001	32	Adam	42	64 x 64
	K Fold							

performance on the SAMM dataset, with validation accuracies of 99.73% and 99.66% and F1-scores of 99.73% and 99.66% for Stratified K-Fold and K-Fold cross-validation, respectively, indicating strong and reliable micro-expression recognition as shown in Table III

TABLE III
METHODOLOGY RESULTS

Proposed Method	Cross Validation Technique	Dataset	Validation Loss	Validation Accuracy	Precision	Recall	F1-Score
Broadbent Attention Layer	Stratified K Fold	SAAM	0.0120	99.72	99.73	99.72	99.72
	K Fold		0.0149	99.65	99.66	99.65	99.65

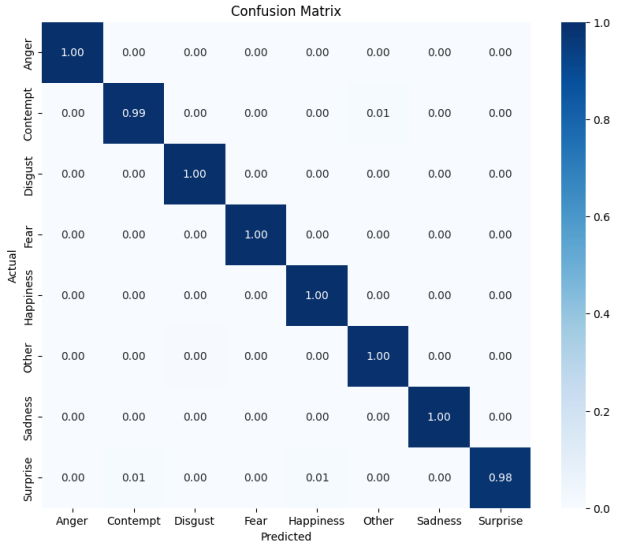

Fig. 2. Confusion Matrix of K-Fold Cross-Validation: Visualizing Model's Emotion Classification Accuracy Across Multiple Categories

Figure 2 shows that the model achieves 100% accuracy for several emotions like Anger, Disgust, Fear, Happiness, and Other. However, there are minor misclassifications between Contempt and Surprise, and between Contempt and Other.

The ROC curve shows near-perfect classification, with all emotions except Surprise having an AUC of 1.00. Surprise achieves 0.99, still indicating excellent performance as shown in figure 3

Proposed model was evaluated using K-fold cross-validation to ensure a robust assessment of its MER capabilities. Each fold was trained and validated on different portions of the dataset to minimize bias and variance in the results. Figure **??**

Authorized licensed use limited to: UNIVERSITY TEKNOLOGI MALAYSIA. Downloaded on January 16,2025 at 04:58:00 UTC from IEEE Xplore. Restrictions apply.

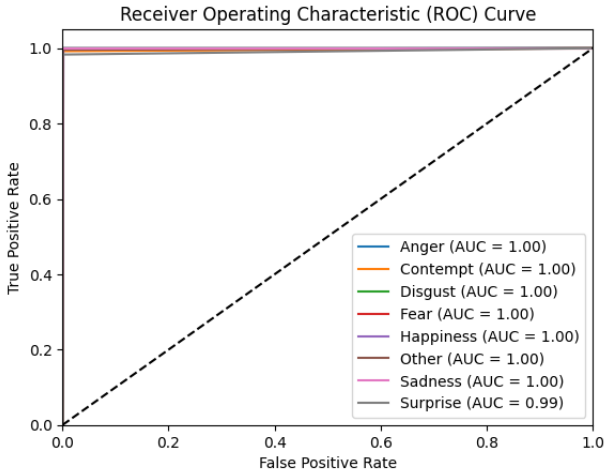

Fig. 3. ROC Curve using K-Fold cross-validation, showing Area Under the Curve (AUC) for different emotion classes

shows validation accuracy and validation loss across multiple folds. The validation accuracy graph shows a general upward trend, with most folds stabilizing near 1.0 after around 10 epochs, though some fluctuations occur. The validation loss graph displays a decrease over time, with spikes in certain folds, but overall, most folds converge to lower loss values as training progresses.

Figure **??** shows training accuracy and loss across folds. Training accuracy quickly rises, stabilizing near 1.0 after about 10 epochs. Training loss sharply decreases, converging to near zero around the same point, with minimal variation across folds.

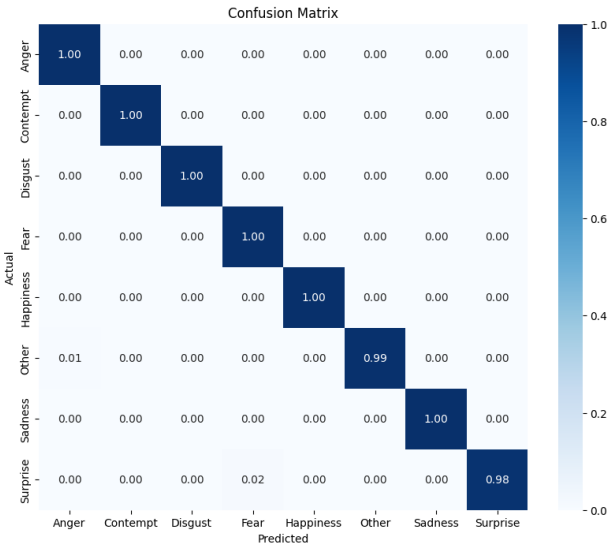

Fig. 4. Confusion Matrix of Stratified K-Fold Cross-Validation: Visualizing Model's Emotion Classification Accuracy Across Multiple Categories

The model performs well in classifying emotions like Fear,and Sadness (all above 0.98), with slight misclassifications in emotions like Other and Surprise as shown in Figure 4. Our Mechanism improve feature focus and generalizability.

Figure **??** depicts validation accuracy and loss across stratified folds. Validation accuracy generally improves and stabilizes near 1.0 after about 10 epochs, though fluctuations occur. Validation loss decreases overall but shows variability, with some spikes, especially in later epochs, before stabilizing at lower values.

Figure **??** shows training accuracy and loss across stratified folds. Training accuracy rises quickly, stabilizing near 1.0 after around 10 epochs. Training loss rapidly decreases, converging close to zero, with minimal variation across all folds. The model demonstrates consistent learning across stratified folds.

Figure **??** shows the model has made predictions for the various facial emotions. The accuracy of 100 percent is shown for categories starting from Contempt to Disgust, Anger, Surprise, Happiness, Sadness, Fear, and Other. With all of the predictions following the labels, it was thus clear that the model was really strong in emotion recognition. The results also indicate how very acute the model is in distinguishing facial expressions and specific subtle micro-expressions.

A. *Quantitative Analysis*

In this section, we analyze the performance of our proposed model featuring the Broadbent Attention Layer and compare it with state-of-the-art methods as mentioned in IVfor emotion recognition. Our evaluation is based on various metrics, including accuracy, precision, recall, and F1-score.

TABLE IV
COMPARISON WITH STATE-OF-THE-ARTS

Model Architecture	**Accuracy**		**Precision**		**Recall**		**F1-Score**	
Proposed Model	***KFold***	***Stratified KFold***	***KFold***	***Stratified KFold***	***KFold***	***Stratified KFold***	***KFold***	***Stratified KFold***
	99.66	***99.73***	***99.66***	***99.73***	***99.66***	***99.73***	***99.66***	***99.73***
GTS-GN [3]	72.06		74.2		73.9		70.9	
MACNN [31]	73.0		72.0		74.0		73.5	
ERFC [21]	94.8		94.5		93.2		90.1	
HTNet(Baseline Method) [38]	81.0		81.22		81.24		81.31	

Table IV compares the proposed model's performance against several state-of-the-art architectures in MER. Proposed model achieves significantly higher accuracy (99.66% and 99.73% for KFold and Stratified KFold, respectively), recall, precision, and F1-score compared to GTS-GN, MTCNN, and HTNet, showcasing its superior capability in recognizing subtle micro-expressions. The results illustrate the effectiveness of the Broadbent Attention Layer method in outperforming existing models.

V. CONCLUSION

This research demonstrates the effectiveness of the Broadbent Attention Layer for micro-expression recognition, achieving near-perfect accuracy on the SAMM dataset with vali-

Authorized licensed use limited to: UNIVERSITY TEKNOLOGI MALAYSIA. Downloaded on January 16,2025 at 04:58:00 UTC from IEEE Xplore. Restrictions apply.

dation accuracies of 99.72% (Stratified K-Fold) and 99.65% (K-Fold), and F1-scores of 99.73 and 99.66. These results highlight the attention mechanism's ability to capture subtle micro-expressions, outperforming many state-of-the-art methods. Model's precision, recall, and F1-score affirm its reliability for applications in fields like behavioral analysis and security.

Future work will extend this model to datasets like CASME and SMIC to assess generalization, and it will be embedded into a web application for practical use in detecting micro-expressions from uploaded images. This tool could benefit sectors like security, psychology, and human-computer interaction, testing the model in real-world settings.

REFERENCES

[1] K. Mohan, A. Seal, O. Krejcar, and A. Yazidi, "FER-net: facial expression recognition using deep neural net," Neural Comput. Appl., vol. 33, no. 15, pp. 9125–9136, 2021.

[2] Z. Zhai, J. Zhao, C. Long, W. Xu, S. He, and H. Zhao, "Feature representation learning with adaptive displacement generation and transformer fusion for micro-expression recognition," in 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023.

[3] J. Wei, W. Peng, G. Lu, Y. Li, J. Yan, and G. Zhao, "Geometric graph representation with learnable graph structure and adaptive AU constraint for micro-expression recognition," IEEE Trans. Affect. Comput., vol. 15, no. 3, pp. 1343–1357, July-Sep 2024.

[4] Z. Wang et al., "Two-level spatio-temporal feature fused two-stream network for micro-expression recognition," Sensors (Basel), vol. 24, no. 5, p. 1574, 2024.

[5] L. Wang, J. Jia, and N. Mao, "Micro-Expression Recognition Based on 2D-3D CNN," in 2020 39th Chinese Control Conference (CCC), 2020.

[6] C. Wu and F. Guo, "TSNN:Three-StreamCombining2Dand3DConvolutional neural network for micro-expression recognition," IEEJ Trans. Electr. Electron. Eng., vol. 16, no. 1, pp. 98–107, 2021.

[7] Y. Li, X. Huang, and G. Zhao, "Micro-expression action unit detection with spatial and channel attention," Neurocomputing, vol. 436, pp. 221–231, 2021.

[8] S.-J. Wang, Y. He, J. Li, and X. Fu, "MESNet: A convolutional neural network for spotting multi-scale micro-expression intervals in long videos," IEEE Trans. Image Process., vol. 30, pp. 3956–3969, 2021.

[9] L. Yao, Y. Wan, H. Ni, and B. Xu, "Action unit classification for facial expression recognition using active learning and SVM, " Multimed," Multimed. Tools Appl, vol. 80, no. 16, pp. 24287–24301, 2021.

[10] Y. Wang et al., "FERV39k: A large-scale multi-scene dataset for facial expression recognition in videos," in 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.

[11] B. Yang, J. Cheng, Y. Yang, B. Zhang, and J. Li, "MERTA: micro-expression recognition with ternary attentions," Multimed. Tools Appl., vol. 80, no. 11, pp. 1–16, 2021.

[12] R. Zhi, C. Zhou, T. Li, S. Liu, and Y. Jin, "Action unit analysis enhanced facial expression recognition by deep neural network evolution," Neurocomputing, vol. 425, pp. 135–148, 2021.

[13] L. Pham, The Huynh Vu, and T. A. Tran, "Facial expression recognition using residual masking network," in 2020 25th International Conference on Pattern Recognition (ICPR), 2021.

[14] Q. Mao, L. Zhou, W. Zheng, X. Shao, and X. Huang, "Region attention and graph embedding network for occlusion objective class-based micro-expression recognition," arXiv [cs.CV], 2021.

[15] Z. Jie, Z. Yuan, S. Jingang, L. Cheng, C. Hongli, and Z. Wenming, "Learning to rank onset-occurring-offset representations for micro-expression recognition," arXiv [cs.CV], 2023.

[16] Z. Wang, K. Zhang, W. Luo, and R. Sankaranarayana, "HTNet for micro-expression recognition," Neurocomputing, vol. 602, no. 128196, p. 128196, 2024.

[17] L. Zhou, Q. Mao, X. Huang, F. Zhang, and Z. Zhang, "Feature refinement: An expression-specific feature learning and fusion method for micro-expression recognition," Pattern Recognit., vol. 122, no. 108275, p. 108275, 2022.

[18] X.-B. Nguyen, C. N. Duong, X. Li, S. Gauch, H.-S. Seo, and K. Luu, "Micron-BERT: BERT-based facial micro-expression recognition," in 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023.

[19] S. Zhao et al., "More is better: A database for spontaneous micro-expression with high frame rates," arXiv [cs.CV], 2023.

[20] S. Zhao et al., "A two-stage 3D CNN based learning method for spontaneous micro-expression recognition," Neurocomputing, vol. 448, pp. 276–289, 2021.

[21] X. Ben et al., "Video-based facial micro-expression analysis: A survey of datasets, features and algorithms," IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 9, pp. 5826–5846, 2022.

[22] H. Li, M. Sui, Z. Zhu, and F. Zhao, "MMNet: Muscle Motion-Guided Network for Micro-Expression Recognition," in Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, 2022.

[23] L. Zhang, X. Hong, O. Arandjelovic, and G. Zhao, "Short and Long Range Relation Based Spatio-Temporal Transformer for Micro-Expression Recognition," IEEE Trans. Affect. Comput., vol. 13, no. 4, pp. 1973–1985, 2022.

[24] S. P. Yadav, "RETRACTED ARTICLE: Emotion recognition model based on facial expressions, " Multimed," Multimed. Tools Appl, vol. 80, no. 17, pp. 26357–26379, 2021.

[25] A. Sepas-Moghaddam, A. Etemad, F. Pereira, and P. L. Correia, "CapsField: Light field-based face and expression recognition in the wild using capsule routing," IEEE Trans. Image Process., vol. 30, pp. 2627–2642, 2021.

[26] H. Li, H. Niu, and F. Zhao, "From Macro to Micro: Boosting micro-expression recognition via pre-training on macro-expression videos," arXiv [cs.CV], 2024.

[27] X. Nie, M. A. Takalkar, M. Duan, H. Zhang, and M. Xu, "GEME: Dual-stream multi-task GEnder-based micro-expression recognition," Neurocomputing, vol. 427, pp. 13–28, 2021.

[28] J. Wang, Y. Tian, Y. Yang, X. Chen, C. Zheng, and W. Qiang, "Meta-auxiliary learning for micro-expression recognition," arXiv [cs.CV], 2024.

[29] M. Verma and S. K. Vipparthi, "Deep insights of learning based micro expression recognition: A perspective on promises, challenges and research needs," 2022.

[30] Y. Liu, W. Wang, C. Feng, H. Zhang, Z. Chen, and Y. Zhan, "Expression snippet transformer for robust video-based facial expression recognition," Pattern Recognit., vol. 138, no. 109368, p. 109368, 2023.

[31] Z. Lai, R. Chen, J. Jia, and Y. Qian, "Real-time micro-expression recognition based on ResNet and atrous convolutions," J. Ambient Intell. Humaniz. Comput., vol. 14, no. 11, pp. 15215–15226, 2023.

[32] M. Aouayeb, W. Hamidouche, C. Soladie, K. Kpalma, and R. Seguier, "Micro-expression recognition from local facial regions," Signal Process," Image Commun, vol. 99, no. 116457, 2021.

[33] L. Luo, J. He, and H. Cai, "The method for micro expression recognition based on improved light-weight CNN," in Lecture Notes in Electrical Engineering, Singapore; Singapore: Springer Nature, 2022, pp. 760–768.

[34] S. C. Ayyalasomayajula, B. Ionescu, and D. Ionescu, "A CNN approach to micro-expressions detection," in 2021 IEEE 15th International Symposium on Applied Computational Intelligence and Informatics (SACI), 2021.

[35] D. Garg and G. K. Verma, "An improved DCNN based facial micro-expression recognition system," in Transactions on Computer Systems and Networks, Singapore: Springer Singapore, 2021, pp. 349–363.

[36] X. Shu, J. Li, L. Shi, and S. Huang, "RES-CapsNet: an improved capsule network for micro-expression recognition," Multimed. Syst., vol. 29, no. 3, pp. 1593–1601, 2023.

[37] A. K. Davison, C. Lansley, N. Costen, K. Tan, and M. H. Yap, "SAMM: A Spontaneous Micro-Facial Movement Dataset," IEEE Trans. Affect. Comput., vol. 9, no. 1, pp. 116–129, 2018.

[38] W. Zhifeng, Z. Kaihao, L. Wenhan, and S. Ramesh, "HTNet for micro-expression recognition," arXiv [cs.CV], 2023.

Authorized licensed use limited to: UNIVERSITY TEKNOLOGI MALAYSIA. Downloaded on January 16,2025 at 04:58:00 UTC from IEEE Xplore. Restrictions apply.