JASTEC

JOURNAL OF APPLIED SCIENCE, TECHNOLOGY AND COMPUTING
e-ISSN: 3036-0250

Vol. 3 No. 1 (2026) 22-31
https://publisher.uthm.edu.my/ojs/index.php/ijie

Advancements and Challenges in Emotion Recognition from Facial Expressions: A Comprehensive Review

Teoh Kuan Yiau[1], Pei-Chun Lin[1]*, Rusma Anieza Ruslan[2]

[1] *Department of Information Engineering and Computer Science, Feng Chia University, No. 100, Wenhwa Road, Seatwen, Taichung 40724, TAIWAN*
[2] *Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, Parit Raja, Batu Pahat, 86400, MALAYSIA*

*Corresponding Author: peiclin@fcu.edu.tw
DOI: https://doi.org/10.30880/jastec.2026.03.01.003

Article Info

Received: 23 January 2026
Accepted: 12 June 2026
Available online: 30 June 2026

Keywords

Emotion recognition, facial expressions, deep learning, data augmentation, human-computer interaction

Abstract

Emotion recognition from facial expressions (FER) is a crucial component of human-computer interaction, enabling machines to comprehend and respond to human emotions. This review examines advancements in FER technologies, highlighting key models such as Deep Residual Networks (ResNet), BReG-NeXt-50, and VGGNet, as well as innovative data augmentation techniques that optimize model performance. Despite significant progress, challenges remain, including class imbalance, generalization issues, and real-time processing constraints. Future research directions are proposed, focusing on the development of hybrid models, dynamic data augmentation, and cross-domain adaptation strategies, as well as ethical considerations and the importance of interdisciplinary collaboration. By addressing these challenges and exploring innovative methodologies, this review aims to advance the development of more robust and effective emotion recognition systems, ultimately facilitating better human-computer interactions across various applications.

1. Introduction

Emotion recognition from facial expressions (FER) is a pivotal aspect of human-computer interaction, enabling machines to understand and respond to human emotions effectively. The significance of FER extends beyond simple emotional assessment; it plays a crucial role in applications ranging from mental health diagnosis to social robotics, where understanding human emotions can enhance the quality of interaction [1]. As computational devices become more integrated into daily life, the ability of these systems to respond to emotional cues becomes increasingly vital, presenting both opportunities and challenges.

Advances in deep learning techniques, particularly convolutional neural networks (CNNs), have significantly improved FER accuracy. State-of-the-art models have demonstrated remarkable performance in controlled environments; however, challenges remain regarding model performance and generalization across different datasets. Variability in lighting, background, and subject demographics can adversely affect model accuracy, leading to performance discrepancies in real-world applications [2]. Furthermore, the tendency of conventional models to overfit to training data poses significant challenges when deployed in diverse scenarios, necessitating the development of more robust architectures.

This is an open access article under the CC BY-NC-SA 4.0 license.

This review outlines the current state of research in FER, emphasizing key models such as ResNet, BReG-NeXt-50, and VGGNet, while also examining innovative data augmentation strategies that enhance model training and robustness. A detailed discussion on the limitations and challenges faced by existing methodologies will provide insight into the complexities of FER. By critically evaluating the current literature and identifying gaps, this review aims to highlight areas that require further investigation, which could lead to improved performance and broader applicability of FER systems.

In recent years, the integration of generative models, particularly Generative Adversarial Networks (GANs), has ushered in a new era of data augmentation for FER [3]. This approach enables the synthesis of diverse facial expressions, enriching training datasets and mitigating class imbalance. The emergence of advanced techniques, such as StarGAN and its enhancements, has shown promising results, revealing the potential for synthetic images to enhance recognition rates and provide more comprehensive training material. However, the practicality of these methods in real-world settings remains an area of active exploration.

As the field progresses, interdisciplinary collaboration becomes increasingly important. The intersection of psychology, computer vision, and artificial intelligence can yield deeper insights into emotional expression and perception. Integrated approaches that combine technological advancements with psychological theories may provide more nuanced interpretations of human emotions, leading to models that not only recognize but also understand emotional complexities. This holistic perspective can pave the way for more empathetic and responsive AI systems, benefiting various sectors ranging from healthcare to entertainment.

This article includes several main sections, including a literature study of emotion recognition from facial expression, data augmentation techniques, Future Direction, and conclusion

2. Literature Review

This section discusses the existing body of work related to emotion recognition from facial expressions.

2.1 Key Frameworks

Recent studies have introduced several influential architectures addressing the complexities of FER. Deep Residual Networks (ResNet) represent a significant innovation in deep learning, primarily designed to combat the degradation problem encountered in very deep networks [4]. This architecture employs the concept of residual learning, where the model learns to predict the residuals or differences between the input and desired output rather than attempting to map the input to the output directly. This approach facilitates better gradient flow during backpropagation, empowering the training of networks with hundreds or even thousands of layers while maintaining high accuracy [5]. The introduction of skip connections allows gradients to bypass certain layers, helping to prevent vanishing gradients and improving stability during training. As a result, ResNet has not only excelled in FER tasks but has also set a benchmark for various other computer vision applications. This influencing subsequent architectures significantly [2]. Fig. 1 illustrates the architecture of ResNet-50, a deep convolutional neural network that uses residual learning to facilitate the training of very deep networks.

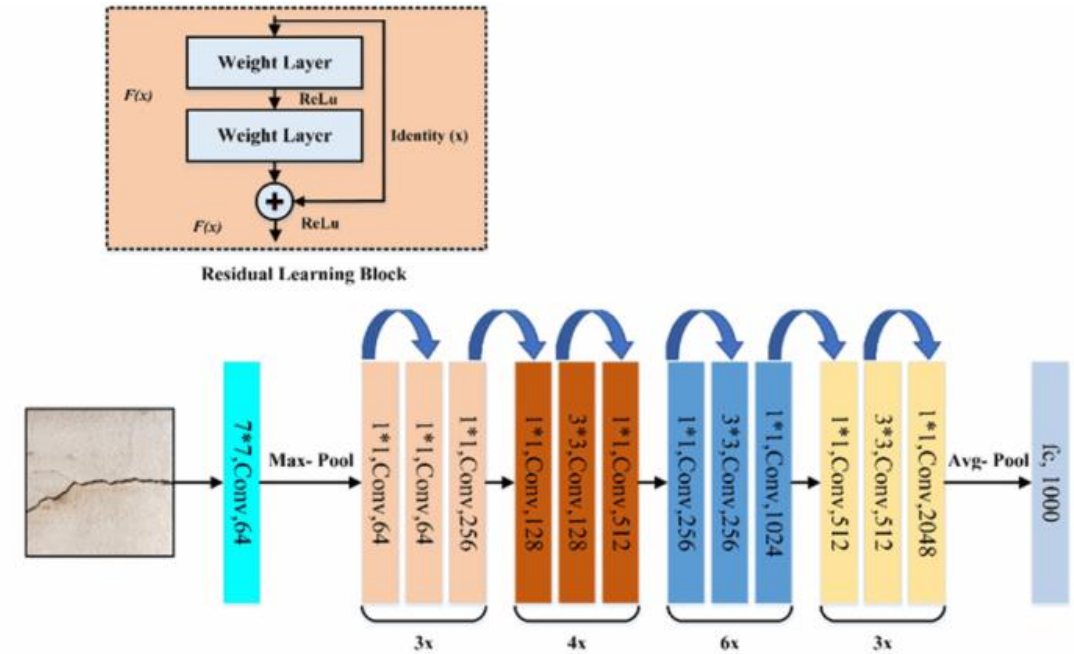


Fig. 1 *ResNet-50 architecture*

J. Appl. Sci. Technol. Comput. Vol. 3 No. 1 (2026) p. 22-31

Building on the foundations laid by ResNet, BReG-NeXt-50 enhances gradient-flow stability by using bounded gradient functions [6]. This architectural improvement specifically addresses the vanishing gradient problem more effectively than traditional ResNet architectures. By incorporating arctangent-based mappings for its shortcut connections, BReG-NeXt-50 retains the crucial advantages of residual learning while ensuring that the gradient flow is well controlled throughout the network. This careful design choice results in a model that achieves superior performance on FER tasks while using significantly fewer parameters than its ResNet counterparts [7]. The efficiency of BReG-NeXt-50 makes it an appealing choice for real-world applications where computational resources are at a premium while still delivering high accuracy in emotion detection. Fig. 2 depicts the network configuration of BReG-NeXt-50, an advanced architecture that extends the principles of ResNet.

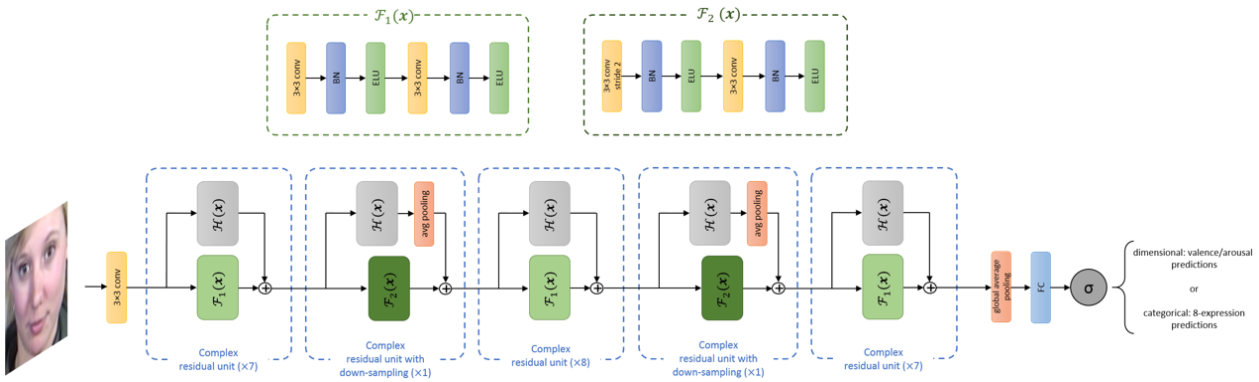


Fig. 2 *Network configuration of BReG-NeXt-50*

VGGNet takes a different approach by focusing on the impact of network depth and filter size on image recognition capabilities [8]. This architecture employs a series of small 3x3 convolutional filters, which have been shown to effectively extract features from images when deployed in a deeper network structure. The design philosophy behind VGGNet underscores that increasing depth, rather than width, can yield superior feature representation and improved performance [9]. Recent optimizations of VGGNet for specific datasets, such as FER2013, have set impressive benchmarks in single-network performance. This emphasizes the critical roles of hyperparameter tuning and data augmentation strategies in achieving high accuracy. VGGNet's findings illustrate the necessity of not only innovative architectural design but also meticulous configuration and enhancement techniques in developing robust models for facial emotion recognition and beyond [10]. Fig. 3 presents the VGGNet architecture, known for its simplicity and depth.

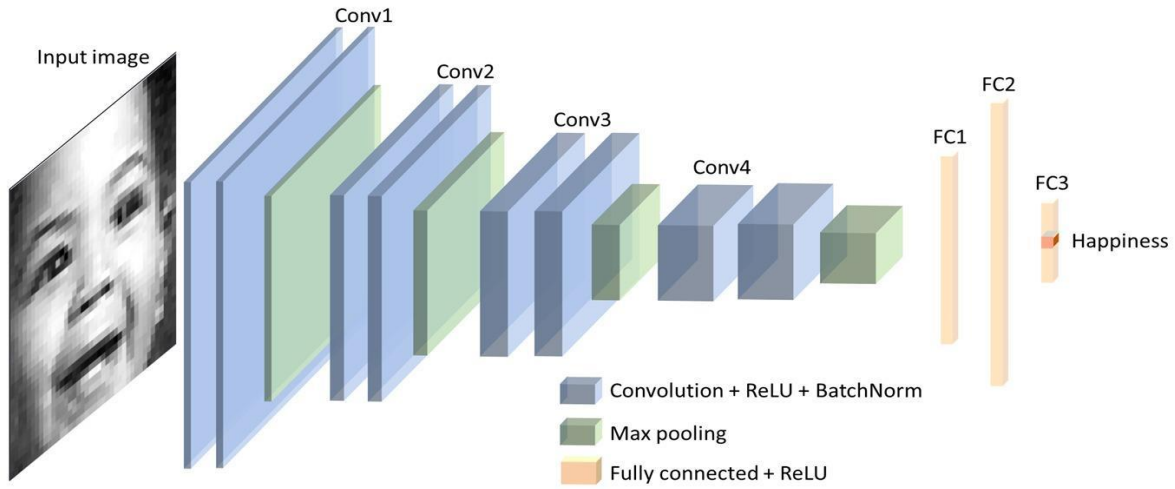


Fig. 3 *VGGNet architecture*

2.2 Data Augmentation Techniques

Data augmentation is vital for improving the robustness of FER models, and traditional geometric transformations remain foundational techniques in this realm. Methods such as rotations, scaling, and flipping are widely used to artificially expand training datasets. It is allowing models to better generalize across various conditions. For instance, rotating images helps the model adapt to different head orientations. On the other hand, scaling can

simulate variations in distance from the camera. Flipping images horizontally addresses variations in left and right facial expressions. It ensures that the model is not biased toward one side of the face. These simple yet effective transformations expose the model to a broader range of facial expressions and variations in lighting, pose, and occlusion, ultimately enhancing its capability to recognize emotions in real-world scenarios.

In recent years, advanced techniques like GANs have revolutionized the landscape of data augmentation for FER [11]. GANs are powerful tools that can synthesize high-quality, diverse datasets, addressing the critical issue of class imbalance often found in FER datasets. By effectively generating new facial expressions, GANs offer an innovative solution for expanding training data and improving model accuracy. Variants like StarGANv2 have taken this a step further by enabling multi-domain facial expression generation, allowing models to learn from a wider array of emotional representations [12]. Additionally, incorporating attention mechanisms within the GAN architecture helps focus on salient facial features, further refining the data generation process. These advancements not only enhance model performance on FER tasks but also foster greater diversity within training datasets, leading to more comprehensive emotion recognition capabilities.

Beyond traditional geometric techniques and GANs, novel augmentation strategies like MixCut are emerging as transformative approaches to data augmentation [13]. MixCut involves creating new training samples by combining attributes from multiple images, allowing the model to learn complex features and relationships between facial expressions. This innovative method creates a richer training environment, leading to improved accuracy compared to conventional techniques that rely solely on basic transformations [14]. By blending different facial attributes, MixCut enables models to experience a wider array of facial expressions and contexts, ultimately enhancing their ability to generalize to new, unseen data. The introduction of these sophisticated augmentation techniques marks a significant evolution in how data augmentation is applied in FER.

Another promising approach to data augmentation is progressive augmentation, in which the intensity of transformations is gradually increased during training [15]. This adaptive strategy allows the model to first learn with simpler transformations before being exposed to more complex variations. By progressively introducing augmentation, models can build a solid foundation of recognition skills before being challenged with more difficult and diverse scenarios. This approach can effectively combat overfitting, as the model learns to adapt to variations incrementally rather than being overwhelmed by high levels of augmentation from the outset [16]. Ultimately, progressive augmentation can significantly enhance model resilience, leading to improved performance on unseen data and more robust emotion detection.

3. Challenges and Limitations

This section explores the challenges and limitations encountered in the existing literature surrounding emotion recognition from facial expressions.

3.1 Class Imbalance

Class imbalance is a prevalent challenge in many machine learning applications, particularly in FER systems. In most datasets, emotions such as happiness and sadness are typically well-represented. Whereas subtle emotions like contempt, disgust, and surprise often have significantly fewer samples [17]. This disparity arises from various factors, including the subjective nature of emotional expression and the difficulty in capturing nuanced emotional states in diverse populations. The result is that models trained on such imbalanced datasets tend to favor the majority classes, leading to biased predictions that neglect the minority classes [18]. This lack of representation diminishes the overall effectiveness of FER systems, as they fail to recognize a critical range of emotional expressions.

The consequences of class imbalance in FER models extend beyond mere performance metrics; they can significantly impact the applicability of these systems in real-world scenarios. For example, in applications such as mental health monitoring or human-computer interaction, an inability to accurately recognize fewer common emotions can lead to misguided interventions or misunderstandings of human responses [19]. Furthermore, this bias can exacerbate existing societal stereotypes and contribute to emotional misinterpretations in sensitive contexts, such as counseling and therapy [20]. Thus, addressing class imbalance is not only crucial for enhancing model accuracy but is also essential for ethical considerations in deploying FER systems in real-life applications.

Several strategies can be employed to mitigate the effects of class imbalance in FER. One common approach is to oversample minority classes to ensure that the model is exposed to sufficient examples during training [21]. This technique can involve duplicating existing samples or generating new ones through data augmentation methods. Conversely, undersampling the majority classes can also help to balance the dataset; however, this method carries the risk of losing valuable information from the overrepresented emotions [21].

In addition to these traditional techniques, the advent of synthetic data generation methods, such as GAN-based approaches, offers innovative solutions [22]. GANs can produce realistic facial expressions for minority emotions, enriching the training dataset without the need for additional real-world data collection. This can help

J. Appl. Sci. Technol. Comput. Vol. 3 No. 1 (2026) p. 22-31

ensure that models receive balanced exposure across various emotional categories, improving overall recognition rates.

Another effective approach to address class imbalance is the use of ensemble methods [23]. By combining multiple models, particularly those trained on different subsets of data or with diverse class distributions, ensemble methods can enhance predictive power for minority classes. This strategy allows for a more robust collective decision-making process, often leading to better performance across all categories, including underrepresented emotions.

Furthermore, cost-sensitive learning techniques can be integrated into the model training process [24]. This involves assigning higher misclassification costs to errors on minority-class examples, encouraging the model to focus on learning these less frequent emotional expressions. By modifying the learning objective to account for class disparities, models can achieve more balanced performance across all classes, enhancing the reliability and applicability of FER systems.

3.2 Generalization

Generalization is a fundamental criterion for evaluating machine learning models, referring to their ability to apply learned knowledge to novel, unseen data. In the context of FER, a model trained on a specific dataset may achieve high accuracy within that dataset but perform poorly when exposed to variations in real-world scenarios [25]. Factors contributing to this discrepancy include differences in lighting conditions, head positions, facial occlusions, and demographic variations such as age, gender, and ethnicity. These elements introduce unpredictability that can exploit weaknesses in a model, leading to suboptimal emotion recognition performance in practice.

One of the most pronounced factors affecting generalization is environmental variability. Facial expressions can vary markedly based on external conditions such as lighting and background. For instance, a model trained in a well-lit environment may struggle in dim settings or bright sunlight, where shadows play across the face, obscuring vital features [26]. Moreover, backgrounds that differ significantly from the training data can distract or mislead the model, further complicating emotion recognition efforts. This sensitivity to environmental factors necessitates the development of more adaptable models that can operate effectively across various conditions.

Demographic differences among individuals also pose a significant challenge for generalization. Facial expressions can be influenced by cultural and social factors, leading to variations in how emotions are conveyed and perceived [27]. For example, certain expressions commonly associated with happiness in one culture may not have the same meaning or intensity in another. Additionally, age-related variations in facial musculature or emotional expression can impact recognition accuracy across different age groups. As a result, FER models trained predominantly on homogeneous datasets may fail to generalize well to diverse populations, underscoring the need for more inclusive training data that encompasses a wide range of demographic variables.

To enhance generalization, researchers are exploring various strategies, including domain adaptation and transfer learning techniques. Domain adaptation involves adjusting a model trained on one domain to perform effectively on another, often using techniques that align feature distributions. For instance, models can be trained in synthetic data and then adapted to recognize expressions in real-world images. Thus, it bridges the gap between controlled training environments and dynamic real-world conditions.

Transfer learning is another approach that allows models pre-trained on large-scale datasets to be fine-tuned on smaller, domain-specific datasets [28]. This approach leverages existing knowledge, enabling the model to adapt more rapidly to novel data while retaining its ability to recognize core features of facial expressions. By transferring learned representations from one domain to another, models can maintain performance even when faced with unseen variations.

In addition to these techniques, continuous learning and data augmentation methods can further improve generalization [29]. Continuous learning enables models to adapt gradually as new data becomes available. This ongoing training process allows models to incorporate new information and adjust to changes in the data distribution over time, keeping them relevant and effective in dynamic environments.

Furthermore, employing advanced data augmentation techniques can simulate a variety of real-world conditions during training. By exposing models to a broader range of variations, data augmentation can help models learn to cope with the complexities they will encounter in practical settings.

3.3 Real-Time Processing

Real-time processing is essential in various applications of FER, including social robotics, interactive gaming, surveillance, and virtual reality. In these contexts, timely and accurate emotion recognition can greatly enhance user interaction and experience. For instance, a social robot that can respond instantaneously to a user's emotional state can create a more engaging and intuitive interaction [30]. Similarly, in gaming environments, recognizing players' emotions in real time can allow for adaptive gameplay, enhancing immersion and enjoyment

[31]. However, achieving this requires balancing accuracy with processing speed. It is making real-time capabilities a paramount consideration.

The primary challenge in real-time FER processing lies in the computational demands of state-of-the-art models. Many advanced deep learning architectures, while boasting impressive accuracy, are often resource-intensive [32]. They require substantial processing power and memory, typically necessitating powerful GPUs and extensive system resources. This can result in latency issues, where the time taken to process and analyze facial expressions slows down the system's response time. In scenarios that necessitate immediate feedback, such delays can significantly diminish user experience and system effectiveness.

In striving for real-time processing, there is often a trade-off between accuracy and speed. Models that are finely tuned for high performance in controlled environments may not be optimized for rapid inference in dynamic, real-world settings [33]. This can lead to challenges in deploying such models in applications where quick decision-making is critical. As a result, achieving a balance between maintaining high accuracy in emotion recognition while ensuring low latency is a key focus of ongoing research.

To address these challenges, researchers are exploring various optimization strategies to enable effective real-time FER. One promising approach is model compression, which reduces model size and complexity through techniques such as pruning, quantization, and knowledge distillation. These methods aim to streamline the model without significantly compromising its performance, making it more suitable for real-time applications.

Additionally, the development of lightweight neural networks tailored for speed and efficiency is gaining traction. Architectures such as MobileNet and EfficientNet are designed to provide a good balance between accuracy and computational efficiency, enabling deployment on devices with limited processing power, such as smartphones and embedded systems [34].

To further enhance real-time processing capabilities, parallel processing and edge computing strategies can be employed. By distributing computation tasks across multiple processors or by using specialized hardware accelerators, such as GPUs or TPUs, the processing burden can be reduced, leading to faster inference times. Similarly, edge computing enables data processing closer to the data source rather than relying on centralized cloud servers, reducing latency and enabling quicker real-time responses.

3.4 Ethical Considerations

One of the most pressing ethical challenges in FER systems is the presence of bias in emotion recognition, particularly concerning race, gender, and cultural differences. Models trained on datasets that predominantly represent certain demographics may fail to accurately recognize or interpret emotions expressed by individuals from underrepresented groups. For example, facial expressions may differ in significance and interpretation across cultures, which could lead to misclassifications and reinforce stereotypes [35]. This bias not only raises concerns about fairness but can also contribute to broader societal injustices, particularly in sensitive applications such as law enforcement or mental health services, where incorrect emotional assessments could have serious consequences.

The implications of biased FER models can be profound. Misinterpretations of emotions can lead to misunderstandings in human interactions, potentially resulting in negative social outcomes. For instance, in customer service or healthcare settings, inaccurate emotion recognition may lead to inappropriate responses that exacerbate rather than alleviate a person's distress. In criminal justice scenarios, biased emotion detection can lead to profiling and discriminatory practices, further entrenching systemic issues within society [36]. As FER systems are increasingly integrated into critical decision-making processes, ensuring equitable performance across diverse populations becomes vital to avoid harming vulnerable groups.

To mitigate biases and promote fairness in FER models, it is essential to establish robust ethical frameworks that guide their development and deployment. Researchers and developers must engage in open discussions about the ethical implications of their work and prioritize integrating fairness assessments throughout the research lifecycle. This includes conducting thorough evaluations of training datasets to ensure they are representative of the diverse populations they aim to serve and employing strategies to reduce bias, such as adaptive training techniques or inclusive design practices.

Another critical ethical consideration is transparency in how FER systems function. Stakeholders, including users and affected individuals, should have insight into how emotion recognition models operate and make predictions [37]. Providing explanations for how a model arrives at its decisions can foster trust and accountability. Techniques such as model interpretability and explainable AI can help clarify the reasoning behind predictions, enabling users to understand potential limitations or biases inherent in the system.

Engaging a broad range of stakeholders, including ethicists, community representatives, and experts from diverse backgrounds, is crucial for ethical FER development. This collaborative approach ensures that the perspectives and needs of various populations are considered and integrated into the modeling and deployment processes. By incorporating diverse viewpoints, developers can better address potential biases and create systems that are more equitable and effective.

J. Appl. Sci. Technol. Comput. Vol. 3 No. 1 (2026) p. 22-31

The ethical landscape surrounding FER systems is dynamic, necessitating a commitment to continuous assessment and responsiveness to emerging issues. As technology evolves and societal norms shift, ongoing evaluations of FER systems' performance and impact are essential to addressing new ethical challenges. Establishing mechanisms for regular monitoring and feedback will help ensure that ethical considerations remain central to the development and deployment of FER technologies.

4. Future Directions

One promising avenue is the development of hybrid models that combine traditional model architectures with contemporary techniques, such as attention mechanisms and transformer architectures. Attention mechanisms can significantly enhance the model's ability to focus on relevant facial features, increasing sensitivity to subtle emotional cues that might otherwise be overlooked. By integrating different approaches, researchers can leverage the unique strengths of each method to create more responsive and accurate FER systems. For instance, combining CNNs with recurrent neural networks (RNNs) may offer benefits in temporal emotion recognition, where understanding sequences of facial expressions over time is critical.

The implementation of dynamic data augmentation techniques presents another exciting opportunity for enhancing FER models. By adapting augmentation strategies based on the model's learning progress, it is possible to tailor the training process more closely to the specific challenges it faces. For example, introducing more aggressive augmentation techniques during early training phases, when models are more prone to overfitting, could promote greater diversity and robustness. Conversely, as the model matures, employing subtler augmentation techniques may further refine its capabilities. This adaptive methodology could improve generalization and the effectiveness of emotion recognition across a broader range of conditions.

Cross-domain adaptation is essential for enhancing the robustness of FER models across different datasets and real-world applications. Many existing FER models are trained on specific datasets that may not capture the full diversity of human emotions encountered in varied environments. Developing methods that allow models to adapt to new, unseen domains is crucial for their practical deployment. Techniques such as domain adversarial training and transfer learning can help bridge the gap between training and testing conditions. Research in this area could focus on mechanisms that facilitate learning from limited data in underrepresented domains, ensuring that models retain high accuracy across diverse settings.

Another area of growth lies in explainable AI (XAI), which aims to make the decision-making processes of FER models more transparent. Understanding why a model interprets certain expressions in specific ways can offer insights into its strengths and limitations, fostering trust among users. XAI techniques can help identify which features influence the model's outputs, guide improvements to the training data, and point to potential biases or weaknesses. By building more interpretable models, researchers can ensure that FER systems are used responsibly and effectively across various applications.

Collaboration with fields such as psychology and neuroscience can provide deeper insights into human emotions and improve model training methodologies. Understanding the intricate dynamics of emotional expressions through interdisciplinary research could yield more sophisticated models that reflect the complexities of human affect. Future studies that incorporate findings from these disciplines can enhance the development of more accurate and nuanced emotion recognition systems.

5. Conclusion

As the field of emotion recognition continues to evolve, it is imperative to address existing challenges while leveraging advancements in neural network architectures and data augmentation techniques. The rapid progress in deep learning has opened new avenues for improving the accuracy and efficiency of FER systems, yet significant obstacles remain, requiring concerted efforts from the research community.

One of the key takeaways from this review is the importance of balancing model complexity with interpretability. Although increasingly sophisticated architecture often improves performance, they may also increase overfitting and limit generalizability. Approaches should therefore improve accuracy while preserving transparency so users can understand and trust the model's decisions. This balance will be vital for the deployment of FER technology in sensitive settings, such as healthcare or security, where misunderstandings can have serious consequences.

Moreover, as FER technology enters various applications, ethical considerations must be prioritized. These applications range from customer service chatbots to autonomous vehicles. Ensuring fairness and minimizing biases will be crucial in fostering public confidence in these systems. Researchers need to be vigilant in examining how their models perform across diverse populations and emotional expressions, actively working to mitigate any biases that may arise. This ethical baseline will not only enhance user acceptance but also improve the overall quality of the technology.

In addition, the potential for cross-disciplinary collaboration cannot be overstated. By engaging with experts in psychology, sociology, and neuroscience, researchers can gain insights into the complexity of human emotions

and expressions. This interdisciplinary approach can lead to the development of more holistic models that reflect the nuanced nature of human affect. The integration of this knowledge can significantly enhance the effectiveness of FER systems, ensuring that they are not merely computational tools but also resonate with human emotional experiences.

Finally, the development of comprehensive, diverse datasets is crucial to the future of FER. Creating more extensive datasets that capture a wide range of emotional expressions across contexts and demographics will provide the foundation needed to train robust models. This effort should also account for real-world variability, including factors such as lighting, angles, and cultural expressions of emotion. Open access to such datasets will encourage cooperation within the research community, drive innovation, and foster advancements in emotion recognition technologies.

Acknowledgement

The authors would like to thank the Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia for its support.

Conflict of Interest

Authors declare that there is no conflict of interests regarding the publication of the paper.

Author Contribution

The authors are responsible for the study conception, research design, data collection, data analysis, result interpretation and manuscript drafting.

References

[1] Spezialetti, M., Placidi, G., & Rossi, S. (2020). Emotion recognition for human-robot interaction: Recent advances and future perspectives. Frontiers in Robotics and AI, 7, 532279.

[2] Kopalidis, T., Solachidis, V., Vretos, N., & Daras, P. (2024). Advances in facial expression recognition: a survey of methods, benchmarks, models, and datasets. Information, 15(3), 135.

[3] Singh, S. (2023). Facial expression recognition using convolutional neural networks (CNNs) and generative adversarial networks (GANs) for data augmentation and image generation (Doctoral dissertation, University of Nevada, Las Vegas).

[4] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778.

[5] Borawar, L., & Kaur, R. (2023, March). ResNet: Solving vanishing gradient in deep networks. In Proceedings of International Conference on Recent Trends in Computing: ICRTC 2022 (pp. 235-247). Singapore: Springer Nature Singapore.

[6] Hasani, B., Negi, P. S., & Mahoor, M. (2020). BReG-NeXt: Facial affect computing using adaptive residual networks with bounded gradient. IEEE Transactions on Affective Computing, 1–1.

[7] Akhand, M. A. H., Roy, S., Siddique, N., Kamal, M. A. S., & Shimamura, T. (2021). Facial emotion recognition using transfer learning in the deep CNN. Electronics, 10(9), 1036.

[8] Dodge, S., & Karam, L. (2016, June). Understanding how image quality affects deep neural networks. In 2016 eighth international conference on quality of multimedia experience (QoMEX) (pp. 1-6). IEEE.

[9] Mohit, M. (2016). The Evolution of Deep Learning: A Performance Analysis of CNNs in Image Recognition.

[10] Bhaidasna, Z. C., Swaminarayan, P. R., & Bhaidasna, H. Z. (2023). Enhancing face recognition with deep learning architectures: a comprehensive review. International journal on recent and innovation trends in computing and communication, 11(9), 164-180.

[11] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, *27*.

[12] Han, B., & Hu, M. (2023). The Facial Expression Data Enhancement Method In-duced by Improved StarGAN V2. Symmetry, 15(4), 956.

[13] Yu, J., Liu, Y., Fan, R., Sun, G. (2024). MixCut:A Data Augmentation Method for Facial Expression Recognition. *arXiv preprint arXiv:2405.10489*.

[14] Yu, H., Liu, A., Li, G., Yang, J., & Zhang, C. (2021). Progressive Diversified Augmentation for General Robustness of DNNs: A Unified Approach. IEEE Transactions on Image Processing, 30, 8955–8967.

[15] Porcu, S., Floris, A., & Atzori, L. (2020). Evaluation of data augmentation techniques for facial expression recognition systems. Electronics, 9(11), 1892.

[16] Fayaz, S., Ahmad Shah, S. Z., ud din, N. M., Gul, N., & Assad, A. (2024). Advancements in data augmentation and transfer learning: A comprehensive survey to address data scarcity challenges. Recent advances in computer science and communications (Formerly: Recent patents on computer science), 17(8), 14-35.

[17] Lavanya, M. S., Arun, V., Tapkire, M., & Shichkina, Y. (2024, December). A Review on Facial Expression Recognition Approaches, Datasets and Technologies. In International Conference on Intelligent Systems (pp. 53-68). Cham: Springer Nature Switzerland.

[18] Huang, C., Li, Y., Loy, C. C., & Tang, X. (2019). Deep imbalanced learning for face recognition and attribute prediction. IEEE transactions on pattern analysis and machine intelligence, 42(11), 2781-2794.

[19] Kaur, M., & Kumar, M. (2024). Facial emotion recognition: A comprehensive review. Expert Systems, 41(10), e13670.

[20] Williams, B. M. (2016). Minding our own biases: Using drama therapeutic tools to identify and challenge assumptions, biases and stereotypes. Drama Therapy Review, 2(1), 9-23.

[21] Dong, Q., Gong, S., & Zhu, X. (2018). Imbalanced deep learning by minority class incremental rectification. IEEE transactions on pattern analysis and machine intelligence, 41(6), 1367-1381.

[22] Kim, J. H., & Won, C. S. (2020, November). Emotion enhancement for facial images using GAN. In 2020 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia) (pp. 1-4). IEEE.

[23] Lawpanom, R., Songpan, W., & Kaewyotha, J. (2024). Advancing facial expression recognition in online learning education using a homogeneous ensemble convolutional neural network approach. Applied Sciences, 14(3), 1156.

[24] Zhang, Y., & Zhou, Z. H. (2009). Cost-sensitive face recognition. IEEE transactions on pattern analysis and machine intelligence, 32(10), 1758-1769.

[25] Liu, J., Fu, Q., Yan, H., & Wang, S. (2025, October). Dual-FER: A Dual-Network Approach to Facial Expression Recognition with Enhanced Generalization. In International Conference on Data Security and Privacy Protection (pp. 3-13). Singapore: Springer Nature Singapore.

[26] Asiedu, L., Adebanji, A. O., Oduro, F., & Mettle, F. O. (2015). Statistical evaluation of face recognition techniques under variable environmental constraints. International Journal of Statistics and Probability, 4(4), 93-111.

[27] Li, R., Quang, H., Filipčíková, M., Xu, Y., Kumfor, F., Spehar, B., & McDonald, S. (2025). Why Do Cultures Affect Facial Emotion Perception? A Systematic Review. Journal of Cross-Cultural Psychology, 00220221251334811.

[28] Angriawan, M. (2023). Transfer learning strategies for fine-tuning pretrained convolutional neural networks in medical imaging. Research Journal of Computer Systems and Engineering, 4(2), 73-88.

[29] Qu, H., Rahmani, H., Xu, L., Williams, B., & Liu, J. (2021). Recent advances of continual learning in computer vision: An overview. arXiv preprint arXiv:2109.11369.

[30] Martínez, A., Belmonte, L. M., García, A. S., Fernández-Caballero, A., & Morales, R. (2021). Facial emotion recognition from an unmanned flying social robot for home care of dependent people. Electronics, 10(7), 868.

[31] Sidharta, A. N., Therry, M. S., Moniaga, J. V., & Chowanda, A. (2024). Emotion-Driven Game Adaptation: A Framework Integrating Facial Expression Recognition and Heart Rate Monitoring for Dynamic Difficulty Adjustment. Procedia Computer Science, 245, 1083-1091.

[32] Selvaganesan, J., Sudharani, B., Shekhar, S. C., Vaishnavi, K., Priyadarsini, K., Raju, K. S., & Rao, T. S. (2024). Enhancing face recognition performance: a comprehensive evaluation of deep learning models and a novel ensemble approach with hyperparameter tuning. Soft Computing, 28(20), 12399-12424.

[33] Lin, Y., Wu, Z., Huang, Q., Liu, X., Yin, B., Hu, J., ... & Wang, Z. (2025, April). Efficient fine-tuning strategies for enhancing face recognition performance in challenging scenarios. In ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1-5). IEEE.

[34] Paul, A. K. (2024). FaceLite: A Real-Time Light-Weight Facemask Detection Using Deep Learning: A Comprehensive Analysis, Opportunities, and Challenges for Edge Computing. Computer Networks and Communications, 83-111.

[35] Blais, C., Linnell, K. J., Caparos, S., & Estéphan, A. (2021). Cultural differences in face recognition and potential underlying mechanisms. Frontiers in Psychology, 12, 627026.

[36] Leslie, D. (2020). Understanding bias in facial recognition technologies. arXiv preprint arXiv:2010.07023.

[37] Barker, D., Tippireddy, M. K. R., Farhan, A., & Ahmed, B. (2025). Ethical Considerations in Emotion Recognition Research. Psychology International, 7(2), 43.