



Facial Expression Recognition: A Review of Advanced Machine Learning Techniques

Fahad Abdu Jibrin*, Agbon E.E, Elvis O, Yau I, Makail S.A, Almustapha M.D, Abdu-Aguye U.F, Usman A.D.

Department of Electronics and Telecommunications Engineering,
Ahmadu Bello University, Zaria, Kaduna State, Nigeria.

ABSTRACT

Facial Expression Recognition (FER) has become essential in diverse domains such as human-computer interaction, healthcare, security, and social robotics. This review paper examines the evolution of FER, highlighting the transition from conventional machine learning algorithms to advanced deep learning techniques. Initially, FER systems heavily depended on handcrafted features and conventional classifiers like Support Vector Machines (SVM) and K-Nearest Neighbors (KNN), using techniques such as Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP). Although foundational, these methods struggled with challenges like lighting, pose, and occlusions. The emergence of deep learning has transformed FER, with Convolutional Neural Networks (CNNs) that automatically learn hierarchical feature representations from raw image data, greatly improving recognition accuracy. Moreover, Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks have been utilized to capture the temporal dynamics of facial expressions. This review also emphasizes the latest deep learning algorithms, including CNNs and LSTMs, as well as the growing role of transformer models with attention mechanisms. Despite these advancements, FER remains challenging due to intra-class variability and contextual influences. Future research directions involve developing more robust algorithms, integrating multimodal data, and addressing ethical concerns such as privacy and bias mitigation. This paper aims to offer a brief overview of FER's current state, guiding future research and applications in this dynamic field.

ARTICLE INFO

Article History

Received: September, 2023

Received in revised form: November, 2023

Accepted: January, 2024

Published online: June, 2024

KEYWORDS

Facial expression recognition, machine learning, deep learning, artificial intelligence

INTRODUCTION

Facial expression recognition (FER) is a crucial field within computer vision and artificial intelligence, dedicated to identifying and interpreting human emotions via facial expressions. FER's importance is evident across multiple domains, such as human-computer interaction, security and surveillance, healthcare, and social robotics [1]. Accurately recognizing and responding to human emotions significantly enhances the interactivity and functionality of these systems, establishing FER as an essential aspect of contemporary technology. Early research in FER primarily depended on handcrafted features and conventional machine learning techniques. These approaches required manual extraction of features such as edges,

textures, and landmarks from facial images, followed by the use of classifiers like Support Vector Machines (SVM) and K-nearest neighbors (KNN) to detect expressions. Techniques such as the Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP) were particularly successful in FER tasks [2,3]. However, these methods often faced challenges with variations in lighting, pose, and occlusions, limiting their effectiveness in real-world applications.

The emergence of deep learning has significantly transformed the field of FER by introducing advanced techniques that automatically extract features from data, thus overcoming many limitations of traditional approaches. Convolutional Neural Networks (CNNs) have been particularly impactful due to

Corresponding author: Fahad Abdu Jibrin

✉ fahad03@gmail.com

Department of Electronics and Telecommunications Engineering, Ahmadu Bello University, Zaria.

© 2024. Faculty of Technology Education. ATBU Bauchi. All rights reserved.

their ability to capture hierarchical patterns in images through layers of convolutions and pooling operations. CNNs have shown exceptional performance in FER, with models like VGGNet and ResNet achieving high accuracy on benchmark datasets [4,5]. These networks are particularly effective at recognizing expressions even in challenging conditions such as varying lighting and complex backgrounds. Recurrent Neural Networks (RNNs), especially Long Short-Term Memory (LSTM) networks, have also advanced FER by tackling the temporal dynamics of facial expressions. Unlike static images, facial expressions comprise sequences of movements that offer richer information for recognition. LSTMs are engineered to capture long-term dependencies in sequential data, making them ideal for tasks requiring an understanding of temporal patterns. Studies have demonstrated that integrating CNNs with LSTMs can improve FER performance by utilizing both spatial and temporal features [6].

Another major advancement in FER is the application of Generative Adversarial Networks (GANs) for data augmentation and enhancing model robustness. GANs comprise two neural networks, a generator, and a discriminator, that compete to produce realistic synthetic data. This functionality is especially valuable in FER, where labeled data is frequently scarce and imbalanced. By generating diverse facial expressions, GANs enrich the training dataset and enable models to generalize unseen data more effectively [7]. Recently, transformer models and attention mechanisms have gained prominence in natural language processing and have been adapted for FER. Utilizing their self-attention mechanism, transformers can capture global dependencies in data, providing a new method for managing facial expressions. Initial research indicates that transformers may surpass traditional RNNs and CNNs in some FER tasks, especially those involving intricate interactions between facial features [8].

Despite these advancements, FER continues to be a challenging issue due to factors like intra-class variability, subtle expressions, and the impact of contextual information. Ongoing research explores hybrid models combining different machine-learning (ML) techniques to improve recognition accuracy and robustness [9]. Furthermore, addressing ethical considerations, such as privacy and bias in FER systems, is

essential as these technologies become more integrated into everyday life [10]. In this review, we thoroughly explore the advanced ML techniques that have influenced the field of FER. We delve into their foundational principles, applications, and the challenges they tackle, offering insights into the future trajectory of FER research. By emphasizing the strengths and weaknesses of these methods, our goal is to assist researchers and practitioners in creating more effective and ethical FER systems.

OVERVIEW OF FACIAL EXPRESSION RECOGNITION

FER has undergone substantial development since its early days, shifting from manual approaches to advanced automated systems. Initially, FER depended on manually designed features and conventional machine learning techniques, such as SVMs and KNN, in conjunction with feature extraction methods like HOG and LBP [2, 3]. Although these early methods laid the groundwork for FER, they faced challenges with lighting variations, pose differences, and occlusions, which limited their practical effectiveness. The advent of deep learning (DL) brought a transformative change to FER by introducing CNNs that automatically learn hierarchical feature representations from raw image data, significantly enhancing recognition accuracy [4]. Subsequent advancements included RNNs and LSTM algorithms, which are capable of capturing the temporal dynamics of facial expressions, thereby improving the recognition of sequential movements. FER has numerous applications across various fields. In human-computer interaction, it facilitates more intuitive and responsive interfaces. In security and surveillance, it helps in detecting individuals' emotional states. In healthcare, FER is used to diagnose mental health conditions and monitor patient well-being. In social robotics, it improves human-robot interaction by enabling robots to understand and react to human emotions. These diverse uses underscore the increasing significance of FER in contemporary technology.

MACHINE LEARNING TECHNIQUES IN FER

Facial expression recognition (FER) initially utilized traditional machine learning techniques, such as SVMs and KNN. These methods depended on manually designed features extracted through techniques like HOG and LBP. Many researchers have conducted several

studies about FER based on ML methods over the last two decades [11-13]. For instance, Abdulrahman et al. [11] introduce a technique for FER that employs Gabor wavelet filters for feature extraction, followed by dimensionality reduction through PCA and LBP. The highest achieved result was an average recognition rate of 90% using the combined Gabor+LBP method on the JAFFE database. However, the approach's effectiveness decreases when dealing with variations in facial expressions and occlusions. The authors of [12] employed PCA and LBP to perform FER with SVM as a classifier, achieving encouraging performance. The best result achieved was a recognition rate of 87% on the JAFFE database. A limitation of this approach is its reduced performance on the MUFEE database, where the recognition rate dropped to 77%, potentially due to head rotation in the images.

Similarly, [12] introduces a method that combines LBP, Oriented FAST and Rotated BRIEF (ORB) features for FER. The best result obtained was an accuracy of 93.2% on the CK+ database. However, the approach struggles with lower accuracy in subject-independent scenarios and varying illumination conditions.

ADVANCED MACHINE LEARNING TECHNIQUES IN FER

Many researchers have adopted advanced ML algorithms to address some of the drawbacks of conventional ML methods, such as the frequent difficulties with changes in lighting, pose, and occlusions that constrained their practical use. Specifically, modern FER utilizes sophisticated methods such as CNNs for hierarchical feature extraction, RNNs and LSTM networks for capturing temporal dynamics, and GANs for data augmentation, thereby improving model robustness and performance [14-15]. For

example, the authors of [14] conducted a comparative analysis between LBP and CNN for FER. The highest result achieved was a 97.32% recognition rate using CNN on the CK+ dataset, which significantly surpassed the performance of LBP, which had a lower accuracy on the same dataset. [15] introduces a hierarchically focused, densely connected CNN incorporating a Spatial Transform Network (STN) for FER. The highest accuracy obtained was 97.44% on the FER-2013 dataset, which notably surpasses the performance of existing methods.

Similarly, Li et al. [16] designed an innovative CNN architecture that integrates attention mechanisms with LBP features to enhance FER. This method attained a top accuracy of 98.52% on the JAFFE dataset. This approach improves recognition performance across various challenging datasets by utilizing attention to emphasize important facial features. The authors of [17] present a novel FER algorithm that integrates HOG-HOF descriptors with VGG-LSTM models. The top result demonstrated an improvement of about 1% over current state-of-the-art methods on the eINTERFACE05 database. This technique successfully captures both spatial and temporal information, thereby enhancing recognition accuracy.

Additionally, the authors of [18] present a visual transformer via feature fusion (VTFF) method for FER under challenging conditions. This technique achieved top-tier performance with accuracies of 88.14% on RAF-DB, 88.81% on FERPlus, and 61.85% on AffectNet. The method adeptly combines CNN features and LBP features with global-local attention mechanisms, enhancing recognition accuracy across different datasets.

Table 1 summarizes the recent and conventional advanced ML techniques employed for FER. References:

Table 1: Summary of conventional and advanced ML techniques in FER

| Reference | Model used | preprocessing | Dataset | Metric | Remark | Year |
|-----------|---|-----------------------------|--|------------------|---|------|
| [11] | Gabor Wavelet Transform, PCA, LBP, K-Nearest Neighbor | Gabor Wavelet Transform | Japanese Female Facial Expression (JAFFE) database | Recognition rate | Gabor+LBP achieved an average recognition rate of 90% | 2014 |
| [12] | Support Vector Machines | PCA, LBP feature extraction | JAFFE, MUFEE | Recognition rate | PCA+SVM achieved 87% on | 2015 |

Corresponding author: Fahad Abdu Jibrin

✉ fahad03@gmail.com

Department of Electronics and Telecommunications Engineering, Ahmadu Bello University, Zaria.

© 2024. Faculty of Technology Education. ATBU Bauchi. All rights reserved.



| | | | | | | |
|------|---|--|--|------------------|--|------|
| | (SVM) | | | | JAFFE and 77% on MUFE | |
| [13] | LBP and ORB features with SVM | Face detection and extraction using Dlib, feature normalization, feature fusion | JAFFE, CK+, MMI | Accuracy | The proposed method outperformed several recent methods, achieving high recognition rates across multiple datasets | 2021 |
| [14] | CNN & LBP | LBP feature extraction, SVM classification | CK+, JAFFE, YALE FACE | Recognition rate | CNN achieved 97.32% on CK+, lowest accuracy 31.82% on YALE FACE | 2020 |
| [15] | Convolutional Neural Network (CNN), Spatial Transform Network (STN) | Illumination normalization using histogram equalization, pose normalization, edge detection, data augmentation | FER-2013, CK+ | Accuracy | Proposed system achieved 97.44% accuracy, outperforming existing systems | 2024 |
| [16] | Attention mechanism-based CNN | Illumination normalization, LBP feature extraction, data augmentation | CK+, JAFFE, FER2013, Oulu-CASIA, NCUFE | Recognition rate | Proposed method outperformed state-of-the-art methods, achieving high recognition rates across multiple datasets | 2020 |
| [17] | HOG-HOF and VGG-LSTM | Face detection, feature extraction, data augmentation | eINTERFACE05 | Accuracy | The fusion method outperformed existing state-of-the-art methods, improving the accuracy by approximately | 2024 |

Corresponding author: Fahad Abdu Jibrin

✉ fajad03@gmail.com

Department of Electronics and Telecommunications Engineering, Ahmadu Bello University, Zaria.

© 2024. Faculty of Technology Education. ATBU Bauchi. All rights reserved.



| | | | | | | |
|------|---|--|-------------------------------|---------------------|---|------|
| [18] | Visual Transformers and Attentional Selective Fusion (VTFF) | Face detection, LBP feature extraction, data augmentation | RAF-DB, FERPlus, AffectNet | Accuracy | 1% Achieved state-of-the-art performance with accuracies of 88.14% on RAF-DB, 88.81% on FERPlus, and 61.85% on AffectNet | 2021 |
| [19] | K-Nearest Neighbors (KNN) and Long Short-Term Memory (LSTM) | LBP feature extraction, face detection, normalization | CK+, Oulu-CASIA | F1-score, precision | The proposed system performs better than other state-of-the-art methods, including those that use deep learning systems | 2024 |
| [20] | Deep CNN with Transfer Learning (TL) | Cropping, resizing, normalization, data augmentation | KDEF, JAFFE | Accuracy | Achieved 96.51% on KDEF and 99.52% on JAFFE using DenseNet-161 with 10-Fold CV | 2021 |
| [21] | Squeeze Vision Transformer | Illumination normalization, facial landmark extraction and data augmentation | CK+, MMI, RAF-DB | Accuracy | Squeeze ViT significantly reduces the number of parameters and computations while improving FER performance | 2022 |
| [22] | Fusion of Attention Mechanism and LBP with CNN (RCL-Net) | Face detection and cropping, resizing, normalization, data augmentation | FER2013, FERPLUS, CK+, RAF-DB | Accuracy | Achieved superior generalization capability and robustness in various datasets | 2023 |
| [23] | ResNet-50 | Face detection | Custom dataset | Accuracy | ResNet-50 | 2021 |

Corresponding author: Fahad Abdu Jibrin

✉ fajad03@gmail.com

Department of Electronics and Telecommunications Engineering, Ahmadu Bello University, Zaria.

© 2024. Faculty of Technology Education. ATBU Bauchi. All rights reserved.

| | | | | | | |
|------|--|--|--------------------------------|----------|---|------|
| | | and extraction, data augmentation | | | achieved the highest accuracy (95.39± 1.41%), outperforming other methods | |
| [24] | Deep Convolutional Neural Network (DCNN) | Cropping, resizing, normalization, data augmentation | Custom dataset with 5 emotions | Accuracy | Achieved an accuracy of 78.04% | 2020 |
| [25] | Attentional Convolutional Network | Face detection, normalization, data augmentation | FER2013, CK+, FERG, JAFFE | Accuracy | Achieved significant improvement over previous models | 2021 |

CHALLENGES AND FUTURE DIRECTIONS

FER encounters numerous challenges, such as managing variations in lighting, pose, and occlusions, which can greatly affect model performance. Moreover, the wide range of human expressions and the subtlety of micro-expressions make achieving high accuracy difficult. Future advancements in FER aim to develop more robust algorithms capable of generalizing across diverse conditions and datasets. Combining multimodal data, like integrating facial expressions with voice and body language, could improve recognition accuracy. Ethical considerations, including privacy and bias mitigation, are also essential as FER technology becomes increasingly prevalent in real-world applications.

CONCLUSION

Facial Expression Recognition (FER) has progressed remarkably from traditional machine learning methods to advanced deep learning models. This review traces the transition from handcrafted feature techniques, such as HOG and LBP, to sophisticated algorithms like CNNs, RNNs, and transformers. Although these advancements have enhanced accuracy and robustness, issues like intra-class variability and contextual influences remain. Future research should aim to develop more resilient algorithms, integrate multimodal data, and address ethical concerns. By continuously innovating and refining these methods, FER can become more

applicable and reliable in various real-world contexts.

REFERENCES

1. Li, Shan, and Weihong Deng. "Deep facial expression recognition: A survey." *IEEE transactions on affective computing* 13.3 (2020): 1195-1215.
2. Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*. Vol. 1. Ieee, 2005.
3. Ojala, Timo, Matti Pietikäinen, and David Harwood. "A comparative study of texture measures with classification based on featured distributions." *Pattern recognition* 29.1 (1996): 51-59.
4. Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556* (2014).
5. He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
6. Donahue, Jeffrey, et al. "Long-term recurrent convolutional networks for visual recognition and description." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.
7. Goodfellow, Ian, et al. "Generative adversarial nets." *Advances in neural information processing systems* 27 (2014).

Corresponding author: Fahad Abdu Jibrin

✉ fahad03@gmail.com

Department of Electronics and Telecommunications Engineering, Ahmadu Bello University, Zaria.

© 2024. Faculty of Technology Education. ATBU Bauchi. All rights reserved.



8. Vaswani, Ashish, et al. "Attention is all you need." *Advances in neural information processing systems* 30 (2017).
9. Umar, Abubakar, et al. "Improved Face Recognition Based on Discrete Bat Algorithm for Feature Selection Scheme.
10. Sariyanidi, Evangelos, Hatice Gunes, and Andrea Cavallaro. "Automatic analysis of facial affect: A survey of registration, representation, and recognition." *IEEE transactions on pattern analysis and machine intelligence* 37.6 (2014): 1113-1133.
11. Abdulrahman, Muzammil, et al. "Gabor wavelet transform based facial expression recognition using PCA and LBP." *2014 22nd signal processing and communications applications conference (SIU)*. IEEE, 2014.
12. Abdulrahman, Muzammil, and Alaa Eleyan. "Facial expression recognition using support vector machines." *2015 23rd signal processing and communications applications conference (SIU)*. IEEE, 2015.
13. Niu, Ben, Zhenxing Gao, and Bingbing Guo. "Facial expression recognition with LBP and ORB features." *Computational Intelligence and Neuroscience* 2021.1 (2021): 8828245.
14. Ravi, Rahul, and S. V. Yadhukrishna. "A face expression recognition using CNN & LBP." *2020 fourth international conference on computing methodologies and communication (ICCMC)*. IEEE, 2020.
15. Jayanthi, P., et al. "An Enhanced Method for Recognition of Facial Expressions using Convolutional Neural Network." *2024 5th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*. IEEE, 2024.
16. Li, Jing, et al. "Attention mechanism-based CNN for facial expression recognition." *Neurocomputing* 411 (2020): 340-350.
17. Chouhayebi, Hajar, et al. "Human Emotion Recognition Based on Spatio-Temporal Facial Features Using HOG-HOF and VGG-LSTM." *Computers* 13.4 (2024): 101.
18. Ma, Fuyan, Bin Sun, and Shutao Li. "Facial expression recognition with visual transformers and attentional selective fusion." *IEEE Transactions on Affective Computing* 14.2 (2021): 1236-1248.
19. Hebri, Dheeraj, et al. "Effective facial expression recognition system using machine learning." *EAI Endorsed Transactions on Internet of Things* 10 (2024).
20. Akhand, M. A. H., et al. "Facial emotion recognition using transfer learning in the deep CNN." *Electronics* 10.9 (2021): 1036.
21. Kim, Sangwon, Jaeyeal Nam, and Byoung Chul Ko. "Facial expression recognition based on squeeze vision transformer." *Sensors* 22.10 (2022): 3729.
22. Liao, Jun, et al. "Facial expression recognition methods in the wild based on fusion feature of attention mechanism and LBP." *Sensors* 23.9 (2023): 4204.
23. Li, Bin, and Dimas Lima. "Facial expression recognition via ResNet-50." *International Journal of Cognitive Computing in Engineering* 2 (2021): 57-64.
24. Pranav, E., et al. "Facial emotion recognition using deep convolutional neural network." *2020 6th International conference on advanced computing and communication Systems (ICACCS)*. IEEE, 2020.
25. Minaee, Shervin, Mehdi Minaei, and Amirali Abdolrashidi. "Deep-emotion: Facial expression recognition using attentional convolutional network." *Sensors* 21.9 (2021): 3046.

Corresponding author: Fahad Abdu Jibrin

✉ fajad03@gmail.com

Department of Electronics and Telecommunications Engineering, Ahmadu Bello University, Zaria.

© 2024. Faculty of Technology Education. ATBU Bauchi. All rights reserved.