Vol.8 No.2 April 2025

Multimedia Research

Received 24 March, Revised 28 April, Accepted 6 May

A Short Review on Multimodal Emotion Recognition Methods

Swarna Mugi S
Department of CSE
Sardar Raja College of Engineering, Tenkasi, Tamil Nadu, India
swarnamugi90@gmail.com

Abstract: The recognition and identification of human emotion states through the combination of several signals, such as text, audio, and facial cues, is known as multimodal emotion recognition (MER). MER aims to identify human emotional states by integrating information from multiple sources, such as speech, facial expressions, text, and physiological signals. In the field of human-computer interaction (HCI), healthcare, and intelligent systems, MER is essential. With advancements in machine learning (ML), deep learning (DL), and the increasing availability of rich multimodal datasets, MER has experienced significant progress. Due to the recent advancements in deep learning (DL) methods and the growing availability of multimodal datasets, the area of MER has seen remarkable development, leading to several important research achievements. There are, however, very few in-depth and targeted evaluations of ML and DL-based MER methods. This survey seeks to offer a review of MER depending on ML and DL techniques. A review that considers 15 articles based on MER is considered in this survey. The review includes a study on miscellaneous aspects. The methods including DL-oriented and ML- oriented approaches for MER are examined in this work. Moreover, the types of multi-modalities taken into account in each of the articles are reviewed. Examination of varied performances and consequently the best attainments are reviewed. Finally, the research gaps on MER are conferred.

Keywords: Multimodal Emotion Recognition; Machine Learning; Speech; Facial Expressions; Accuracy.

Nomenclature

Abbreviation	Description
Af-CAN	Attention-based Fusion Contextual Attention Network
AM-DSCNN	Attention Mechanism-Depthwise Separable Convolutional Neural Networks
AI	Artificial Intelligence
BiGRU	Bidirectional Gate Recurrent Unit
BDAE	Bimodal DAE
CA-MDER	Context-Aware Multi-Modal Driver Angry Emotion Recognition
CNN	Convolutional Neural Network
DT	Decision Tree
DAE	Deep Automated Encoder
DNN	Deep Neural Network
DL	Deep Learning
e2e	end-to-end
EEG	Electroencephalography
FF	Feature Fusion
FC	Fully Connected
FE	Facial Expression
FLF	Feature-Level Fusion
GRU	Gated Recurrent Unit
GF	Gated Fusion
GNN	Ghostnet Neural Network
GSR	Galvanic Skin Response
HCI	Human-Computer Interface
HF-CNN	Hierarchical Fusion Convolutional Neural Network
ISSA	Improved Sparrow Search Algorithm
LSTM	Long Short-Term Memory
LFCNN	Lightweight Fully Convolutional Neural Network
MHA	Multi-Head self-Attention
MI	Mutual information
MDAT	Multimodal Dual Attention Transformer
MER	Multimodal Emotion Recognition
ML	Machine Learning

https://doi.org/10.46253/j.mr.v8i2.a4

MFCC	Mel-Frequency Cepstral Coefficient
NN	Neural Network
PPS	Peripheral Physiological Signals
RF	Random Forest
ResGCN	Residual Graph Convolutional Network
SVM	Support Vector Machines
SCME	Sparse Cross-Modal Encoder
T&AF	Transformers and Attention-based Fusion
V2V	Vehicle-To-Vehicle
tLSTM	tree-like LSTM

1. Introduction

Emotion recognition is a crucial field of study as it allows computers to effectively understand human emotions and respond intelligently to human needs. Recognizing human emotions is essential for enabling machines to interact more naturally and effectively with users. Emotions may have a significant impact on students' well-being and performance in educational environments [1]. Teachers can better understand the student's academic achievement and general well-being by using emotion recognition technology to track their emotional states. In the healthcare industry, emotion recognition has huge promise, since technology can help physicians better comprehend their patients' emotional states, allowing for the delivery of customized and specialized medical care. By identifying customers' emotional requirements and providing more individualized services, emotion recognition may also improve intelligent customer care systems. Due to its potential to completely transform the way people and computers interact, emotion detection systems have become a prominent topic of AI research [2].

The recognition of individual modalities, such as text emotion recognition, speech emotion recognition, and face expression recognition, was the main focus of initial emotion recognition research. Early research in this domain primarily focused on unimodal systems, such as speech or facial expression analysis. However, these approaches often struggle with limited accuracy due to data noise and modality-specific limitations. However, researchers are increasingly using multiple modalities to make emotional assessments due to the limited preciseness of single modalities, which suffer from inadequate data and noise. To overcome the limits of single modalities, MER is focused more nowadays. Furthermore, MER can extract the most discriminative and informative characteristics from each modality by using the most MI to measure the reliance between several modality features [3]. This method improves the model's ability to distinguish between different emotions. MER has drawn the attention of numerous researchers to provide more comprehensive and precise info for emotional assessment and to greatly enhance the performance of mental judgments. Its goal is to integrate data from multiple modalities that can support and enhance one another.

Currently, DL-based MER has become one of the most popular study areas in the area of emotion recognition due to the quick advancement of DL techniques. The design of certain networks and related loss functions plays a major role in MER. This adaptability implies that DL methods may be skilfully tailored to capture the relations between different modalities, extracting emotional information-rich features for precise emotion recognition. DL approaches are being used by more and more researchers for MER, with encouraging results. The use of advanced architectures, including transformers and attention mechanisms, is allowing researchers to better model temporal and cross-modal dependencies in emotional data. A thorough analysis of the most updated approaches and a prediction of future areas of study are now essential due to this growing interest [4].

A review is performed that considers 15 articles based on MER with the below contributions.

- Reviews different methods including DL-oriented and ML- oriented approaches adopted for MER.
- Reviews the types of multi-modalities like video, audio, EEG signals, etc. that are taken into account in each of the articles.
- Analyses varied performances and consequently, the best attainments are reviewed.

The overview of MER is explained in Section II. Explanation of DL/ML based MER methods is explained in section III. A review of MER methods and types of multi-modalities is examined in Section IV. The best accomplishments and performances are examined in section V. Research issues are explained in section VI and section VII provides the conclusion.

2. Overview of Multimodal Emotion Recognition

MER is the process of identifying and categorizing human emotions by integrating various signals from voice, text, and facial cues. It includes research on HCI, AI, security, and medical. It is possible to extract

emotions from a variety of sources, including physiological signals (such as EEG and ECG) and physical information (such as text, audio, and video). The fusion of modalities is the characteristic that sets unimodal conduct apart from multimodal or bimodal behavior. Numerous attempts to create DL-based schemes with excellent outcomes can be found in the literature as a result of DL's significant success in emotion identification and classification, especially using CNN. While some researches use an unimodal method, others deploy bimodal or multimodal approaches.

Fig. 1 shows a general pipeline of multimodal emotion recognition, illustrating the flow from input modalities (text, audio, video, EEG) through feature extraction and fusion (early, late, hybrid), to final emotion classification.

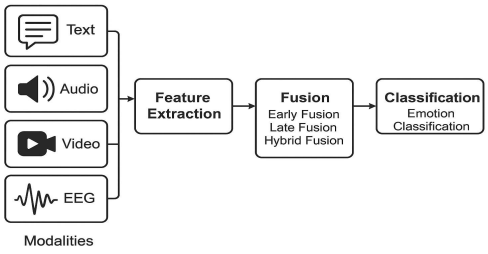

Fig. 1*. A generalized pipeline of multimodal emotion recognition (MER).*

3. Review on DL/ML based MER Methods

3.1 Methods based on DL

The Af-CAN is a MER technique that was suggested by Xue et al. [1] in 2024. To guarantee the thoroughness and precision of emotion detection, Af-CAN was built with a multimodal FF system that can gather emotion-relevant data from various information sources and merge these elements using sophisticated attention processes. This approach enhances the ability to process complex emotional cues and improves emotion recognition performance. Additionally, this framework presented a unique context modeling unit that can monitor the changes in emotional states during the discussion as well as the reciprocal impact of emotions among speakers.

Yucel et al. [2] presented a new emotion identification system in 2020 that used a variety of modalities, including EEG, FE, and GSR. This approach used a hybrid fusion approach and achieved an accuracy of up to 81.2%. It was noted for its robustness in detecting emotions under varying conditions. When it comes to natural deceitful facial expressions, the model that was being shown was especially helpful in identifying the appropriate emotional state. This research outperformed extant approaches regarding recognition accuracy.

In 2024, Syed et al. [3] suggested a MDAT framework to enhance cross-language MER. The model employed pre-trained algorithms for extracting features from multimodalities and was supplied with dual attention processes to identify intricate connections across multiple modalities and languages to achieve superior cross-language MER. This method has shown promise in multilingual contexts, where emotional expression may differ significantly across languages. To increase the accuracy of emotion categorization, this model further makes use of a transformer encoder layer.

In 2021, Anamaria et al. [4] created TA-AVN, an e2e NN framework that combined both audio and video information in an autonomous environment to determine a subject's emotional state. Real-time processing was achieved by employing basic CNNs to extract the feature descriptor for video and audio representations. The framework proved especially useful in dynamic environments where quick emotional state detection is required. Gathering annotated training data was still a significant barrier regarding the time and skill needed to train emotion detection algorithms. This issue was resolved by the suggested method, which offered a natural augmentation methodology that enabled obtaining high levels of precision even with a small quantity of annotated training data.

The multimodal merging of text and speech for emotion identification was the main emphasis of Yanan Shang and Tianqi [5] in 2024. A characteristic of speech emotion was a 39-dimensional MFCC. The attribute used for text emotion was a 300-multidimensional word vector that was produced using the Glove method. These multidimensional representations allowed the system to capture complex emotional nuances. In DL, the Bi-GRU approach was introduced to extract deep features. The ISSA-BiGRU-MHA approach for emotion identification was then created by combining it with the ISSA and the MHA mechanism.

In 2025, Songül and Fatma [6] used sophisticated models including Transformer, LSTM, and GRU to combine EEG signals with FE to increase the understanding of MER. The results highlighted how multimodal methods provide notable advantages over conventional unimodal techniques by utilizing a multi-class classification framework. This approach demonstrated how combining modalities could increase resilience in detecting subtle emotional shifts. To improve the accuracy and resilience of emotion identification algorithms, this work offered an approach that captured intricate neural processes. These findings have significant applications, demonstrating how the drawbacks of single-modality systems may be addressed by combining many data sources with sophisticated models.

Jiahui et al. [7] presented Deep-Emotion, a DL-based MER, in 2023. It can adaptively incorporate the most discriminating elements from speech, FE, and EEGs to enhance the MER's performance. The method also provided real-time adaptation, adjusting to varying input conditions efficiently. In particular, the EEG, face, and speech branches were the three branches that make up the suggested Deep-Emotion architecture. The facial branch, in turn, employed the enhanced GNN that was suggested for feature extraction. This network successfully mitigated the overfitting issue during training and enhanced the classification accuracy in comparison to the original GhostNet network. For EEG, a tLSTM framework was suggested that can combine multi-stage information. Ultimately, the decision-level fusion approach was used to combine the recognition outcomes of the three aforementioned modes, leading to a more thorough and precise performance.

By integrating the BERT model with heterogeneous information based on audio, language, and visual modalities, Sanghyun et al. [8] introduced a novel MER technique in 2021 that enhanced the BERT model for emotion recognition. In particular, as the auditory and visual modalities have different characteristics, the BERT model was enhanced. This integration of multiple modalities ensured more accurate emotion detection in diverse environments. Using the newly suggested transformer architecture, the "Self-Multi-Attention Fusion module, multi-attention fusion module, and Video Fusion module" were presented, which were attention-based multimodal fusion mechanisms. How to best integrate fine-grained visual and aural feature representations into a shared embedding was investigated while integrating a trained BERT algorithm that includes fine-tuning modalities.

Shuai et al. [9] presented a DL model for MER in 2023 that combined FE and EEG inputs to produce a superior classification impact. First, the face characteristics were extracted from the FE using a CNN that had already been trained. To extract more important face frame elements, the attention process was then added. This allowed the model to focus more on the most crucial features of emotion recognition. After that, CNNs were utilized to extract spatial characteristics from the original EEG signals. After FLF, the model used for emotion recognition received the fused features of EEG and FE data and generated the recognized outputs.

In 2020, Shamane et al. [10] used features taken from separately pre-trained SSL models to represent three input modalities: text, audio (voice), and vision. A unique T&AF method was provided that can merge multimodal SSL information and obtain results for MER tasks, taking into account the high dimensionality of SSL features. This method was able to maintain high accuracy while reducing the computational load. The work was assessed and it was demonstrated that the model was reliable and performed better than the extant methods on four datasets.

To get the possible information in the data for MER, Yong et al. [11] proposed a HF-CNN model in 2021. To create the final feature vector, it builds various network hierarchical structures, extracts multiscale features, and uses FLF to combine the global features created by integrating weights with manually derived statistical features. The multiscale feature extraction proved especially useful for distinguishing between subtle emotional variations. The performance of the suggested model was assessed in this research by binary classification studies.

Mohammed [12] introduced a new MER system in 2024 that used DCNN to identify emotions in text, audio, and video data. Happy, angry, sad, scared, disgusted, surprised, and neutral emotions may all be identified by the system. The ability to identify a wide range of emotions made the system particularly versatile across applications. The system was trained and tested using three datasets, one for all of the 3 input types. According to the findings, the identification accuracy rate for audio, video, and text was 100%, 69%, and 64%, respectively. The recorded accuracy percentage was 80% when using the decision-level fusion. These outcomes demonstrated how well the machine can identify human emotions.

3.2 Methods based on ML

A speed-optimized MER framework for text and speech emotion detection was presented by Lin et al. [13] in 2024. An effective RoBERTa pre-trained network was utilized as the text feature extractor in the feature extraction section, while a compact ResGCN was chosen as the speech feature extractor. These two feature types were then fused using a SCME, which was suggested as a complexity-optimized technique. This approach significantly reduced processing time while maintaining high accuracy. Lastly, several findings

were weighted using a new GF module before being sent into a FC layer for recognition. This approach has a reasonable training and inference time and outperformed the stated approaches regarding accuracy.

To solve the primary problems in MER tasks, TONGQIANG et al. [14] developed a CA-MDER approach in 2024. These included the incapacity to record driving behavior data that reflected V2V contact, individual variances among drivers, and variations in feelings across various driving settings. The integration of driving data made the model particularly applicable to in-vehicle emotion detection. On utilizing facial, verbal, and driving status information, the technique performed multi-modal rage emotion detection utilizing AM-DSCNN, an enhanced SVM, and RF models.

Hongli et al. [15] presented a DAE-based expression-EEG interaction MER technique in 2020. The first approach used for feature selection was the DT. The FE group for test samples was then determined by analyzing the solution vector values in light of the FE characteristics identified via sparse representation. The EEG and facial expression data were then fused using the BDAE. This made the model highly adaptable in dealing with real-time emotional states. The BDAE's third layer gathers characteristics for supervised learning training. Lastly, the classification process was done using the LIBSVM classifier. The outcomes demonstrated that high-level emotion-related characteristics in EEG and FE data could be successfully extracted and integrated using the suggested strategy.

4. Review on MER Methods and Types of Multi-Modalities

4.1 Review of MER Methods

The various MER approaches used in reviewed works are specified in Fig. 1. Here, the MER approaches are categorized into 2 sets i.e., ML- oriented and DL-oriented. The ML-based methods include CNN, LSTM, Bi-GRU, GhostNet, and so on. Initially, Af-CAN was used for multimodal emotion recognition in articles [1], while, another DL-based method called MDAT was deployed in [3] for MER. The most widely deployed DL-based method, CNN is employed for multimodal emotion recognition in articles [2] [4] [9] and [11]. The deep CNN is used for MER in [12]. The BiGRU-based DL method is used for MER in [5] and [6]. LSTM-based DL method and Transformer models are also deployed in [6] along with GRU as ensemble models. Ensemble classifiers including GhostNet, LFCNN, and LSTM are deployed for MER in [7]. The BERT-based method is utilized for MER in [8] and the transformer-based DL model is utilized in [10] for MER. Further, ML is also found to offer promising outcomes on MER. The ML-based SCME approach was adopted for MER in [13], while, article [14] deployed ensemble classifiers including AM-DSCNN, SVM, RF. An enhanced version of SVM termed LIBSVM is adopted for MER in [15].

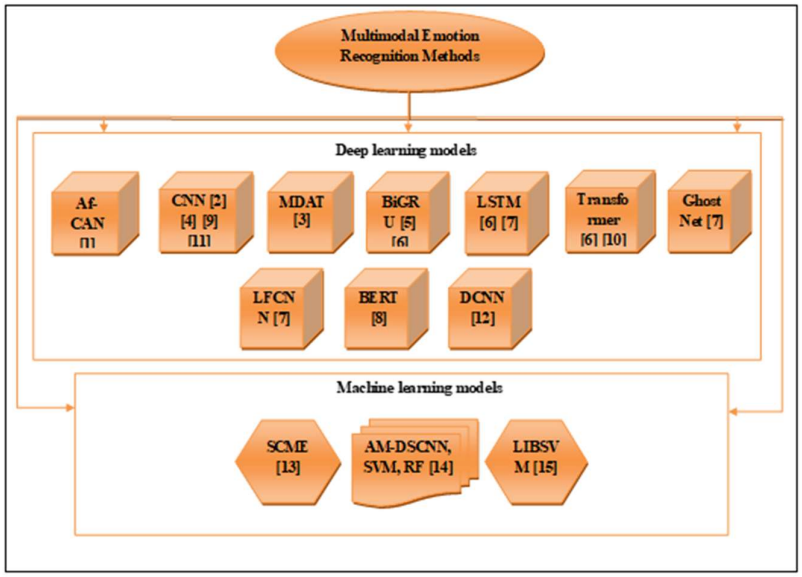

Fig. 2. *Analysis of diverse MER Methods*

Fig. 2 illustrates the analysis of the diverse MER methods.

4.2 Types of Multimodalities

The diverse kinds of multimodalities like text, speech, audio, EEG, and so on, which are adopted in reviewed articles are specified in Table I. Here, text was used as a modality in articles like [1] [3] [5] [10]

[12] and [13]. Voice was used as a modality in [1] and [14]. Video was used as a modality in [1] [4] [8] [10] and [12]. Facial Expressions were used as a modality in [2] [6] [7] [9] [14] and [15]. GSR was used as a modality in [2]. EEG was used as a modality in [2] [6] [7] [9] [11] and [15]. The audio was used as a modality in [3] [4] [8] [10] and [12]. Speech was used as a modality in [5] [7] and [13]. Language was used as a modality in [8]. Vision was used as a modality in [10]. Peripheral Physiological Signals were used as a modality in [11]. Vocal and driving were used as a modality in [14].

Table 1. *Types of Multimodalities Adopted in Reviewed Articles*

Citations	Considered Multimodalities
[1]	Text, Voice, and Video
[2]	Facial Expressions, GSR, and EEG.
[3]	Audio and Text
[4]	Audio and Video
[5]	Speech and Text
[6]	EEG Signals with Facial Expressions
[7]	Facial Expressions, Speech, and EEG
[8]	Language, Audio, and Visual
[9]	EEG Signals and Facial Expressions
[10]	Text, Audio, and Vision
[11]	EEG Signals and PPS
[12]	Audio, Video, and Text
[13]	Speech and Text
[14]	Facial Expressions, Vocal and Driving
[15]	EEG and Facial Expression

5. Review on performances and Best Accomplishments

Table II shows the measures used for analysis in considered articles. The widely adopted measure for analysis includes accuracy and F1-Score. Almost, 11 works considered accuracy as a metric for analyzing the performance of MER. The F1-score measure was analyzed in [1] [8] [10] [14] and [15], i.e. around 5 works have adopted F1-score for analyzing MER. The unweighted accuracy and weighted accuracy were computed in 2 articles each. Articles [3] and [5] considered unweighted accuracy measure, while, articles [5] and [13] considered weighted accuracy measure. The loss was considered as a metric for analyzing the performance of the recognition classifier in [4] and [12]. MAE and correlation were examined in2 articles each. Articles [8] and [10] considered correlation measures for examining the betterment of MER. The running time was examined in [15], which computes the time taken for executing emotion recognition. Certain other metrics like computing efficiency, ablation studies, and confusion matrix were also considered in adopted works, however, the metrics in Table II contribute more.

Table 2. *Analysis of performance measures*

Citations	F1 score	Accuracy	Unweighted accuracy	Loss	Weighted accuracy	MAE	Correlation	Running time	Others
[1]	✓								✓
[2]		✓							✓
[3]			✓						✓
[4]		✓		✓					✓
[5]			✓		✓				✓
[6]		✓							✓
[7]		✓							✓
[8]	✓	✓				✓	✓		
[9]		✓							
[10]	✓	✓				✓	✓		
[11]		✓							
[12]		✓		✓					✓
[13]					✓				✓
[14]	✓	✓							
[15]	✓	✓						✓	✓

5.1 Best Performances

Among the performances, the best ones are reviewed in this section. Fig. 3 and Fig. 4 display the best performances accomplished for accuracy and F1-scores respectively. The major works have analyzed accuracy, which has attained the best performance in many of the articles. Particularly, a higher accuracy

of 100% is accomplished by DCNN in [12] and subsequently, GhostNet, LFCNN, and LSTM [7] have gained a higher accuracy of 98.27%. Further, a higher F1-score of 94% is accomplished by BERT in [8]. Following BERT, LIBSVM in [15] accomplished a higher F1-score value of 92%. A higher weighted accuracy of 82.4% is accomplished by SCME in [13], whereas, a high unweighted accuracy of 90% is accomplished by MDAT in [3].

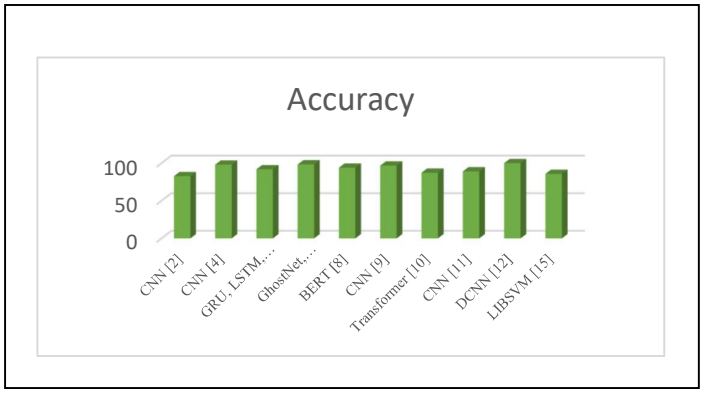

Fig. 3. *Accuracy analysis*

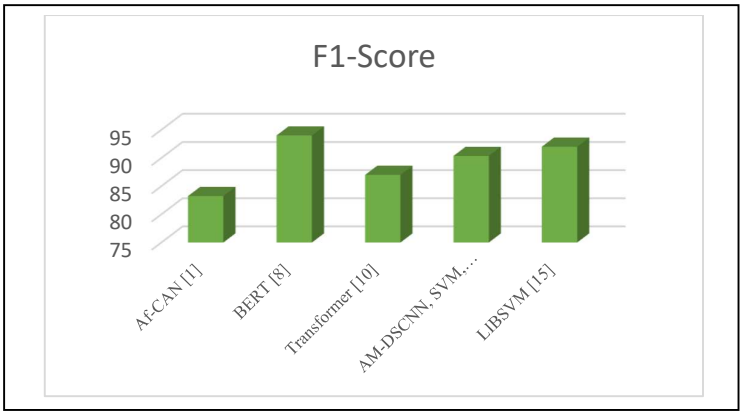

Fig. 4. *F1-score analysis*

6. Research Gaps

Humans use emotion to communicate their requirements, desires, interests, and worries. Emotions serve as a means of communication amongst several living species. In the age of AI, computers are becoming an increasingly significant aspect of our daily lives. However, to respond to the interactions more effectively, they must be able to identify human emotions. MER is the process of identifying and categorizing human emotions by fusing various signals from voice, text, and facial cues. MER faces several key challenges, including the effective fusion of diverse data types like audio, visual, and physiological signals, each with its noise, sampling rates, and feature characteristics. Handling missing or incomplete modalities during training and inference is difficult, especially in real-world settings. Moreover, emotional expressions are highly subjective, culturally dependent, and context-sensitive, making consistent labeling and ground truth creation a major hurdle. The scarcity of large, well-annotated multimodal datasets further limits the development of robust models. Finally, achieving real-time performance while maintaining accuracy and ensuring model interpretability remains an ongoing challenge, particularly for deployment in practical applications like healthcare, education, and HCI.

7. Conclusion

A review was performed that considered 15 articles based on MER with the below analysis.

- Different methods including DL-oriented and ML- oriented approaches adopted for MER were reviewed in this work.

- In addition, the types of multi-modalities like video, audio, EEG signals, etc. that are taken into account in each of the articles for MER were reviewed.
- The varied performances and consequently the best attainments for better MER were reviewed.
- Lastly, the research issues on MER were conferred.

In conclusion, this review highlights the diverse methodologies and multimodal approaches used in emotion recognition, providing valuable insights into the strengths, challenges, and future directions for improving MER systems. The integration of different modalities and advancements in deep learning and machine learning continue to push the boundaries of emotion detection, while also presenting opportunities for addressing current research challenges.

Compliance with Ethical Standards

Conflicts of interest: Authors declared that they have no conflict of interest.

Human participants: The conducted research follows the ethical standards and the authors ensured that they have not conducted any studies with human participants or animals.

References

[1] X. Zhang, M. Wang, X. Zeng, and X. Zhuang, "Af-CAN: Multimodal Emotion Recognition Method Based on Situational Attention Mechanism," IEEE Access, vol. 13, pp. 44858 - 44871, 44858 - 44871.
[2] Y. CIMTAY, E. EKMEKCIOGLU, and SEYMACAGLAR-OZHAN, "Cross-Subject Multimodal Emotion Recognition Based on Hybrid Fusion," IEEE Access, vol. 8, 2020.
[3] S. L. Q. SYED AUN MUHAMMADZAIDI, "Enhancing Cross-Language Multimodal Emotion Recognition With Dual Attention Transformers," IEEE Open Journal of the Computer Society, vol. 5, pp. 684 - 693, 2024.
[4] A. B. N.-C. R. L.-C. D. ANAMARIA RADOI, "An End-To-End Emotion Recognition Framework Based on Temporal Aggregation of Multimodal Information," IEEE Access, vol. 9, 2021.
[5] Y. Shang and T. Fu, " Multimodal fusion: A study on speech-text emotion recognition with the integration of deep learning," Intelligent Systems with Applications, vol. 24, 2024.
[6] S. E. Güler and F. P. Akbulut, "Multimodal Emotion Recognition: Emotion Classification Through the Integration of EEG and Facial Expressions," IEEE Access, vol. 13, pp. 24587 - 24603, 2025.
[7] J. Pan, W. Fang, Z. Zhang, B. Chen, Z. Zhang, and S. Wang, " Multimodal Emotion Recognition Based on Facial Expressions, Speech, and EEG," IEEE Open Journal of Engineering in Medicine and Biology, vol. 5, pp. 396 - 403, 2023.
[8] S. Lee, D. Han and H. Ko, "Multimodal Emotion Recognition Fusion Analysis Adapting BERT With Heterogeneous Feature Unification," IEEE Access, vol. 9, pp. 94557 - 94572, 2021.
[9] S. Wang, J. Qu, Y. Zhang and Y. Zhang, "Multimodal Emotion Recognition From EEG Signals and Facial Expressions," IEEE Access, vol. 11, pp. 33061 - 33068, 2023.
[10] S. Siriwardhana, T. Kaluarachchi, M. Billinghurst and S. Nanayakkara, "Multimodal Emotion Recognition With Transformer-Based Self Supervised Feature Fusion," IEEE Access, vol. 8, pp. 176274 - 176285, 2020.
[11] Y. Zhang, C. Cheng and Y. Zhang, "Multimodal Emotion Recognition Using a Hierarchical Fusion Convolutional Neural Network," IEEE Access, vol. 9, pp. 7943 - 7951, 2021.
[12] M. A. Almulla, "A multimodal emotion recognition system using deep convolution neural networks," Journal of Engineering Research, 2024.
[13] Y. Z. Y. C. B. W. X. S. Lin Cui, "A high-speed inference architecture for multimodal emotion recognition based on sparse cross-modal encoder," Journal of King Saud University- Computer and Information Sciences, vol. 36, no. 5, 2025.
[14] T. Ding, K. Zhang, S. Gao, X. Miao and J. Xi, "A Multimodal Driver Anger Recognition Method Based on Context-Awareness," IEEE Access, vol. 12, pp. 118533 - 118550, 2024.
[15] H. Zhang, "Expression-EEG Based Collaborative Multimodal Emotion Recognition Using Deep AutoEncoder," IEEE Access, vol. 8, pp. 164130 - 164143, 2020.