ijamrsd.com
JAMRSD
ISSN: 2583-6404
Journal of Advanced Multidisciplinary Research Studies and Development
Peer Reviewed Referred Journal Bi-monthly Open Access

ADVANCEMENTS AND CHALLENGES IN EMOTION RECOGNITION: A COMPREHENSIVE REVIEW OF ARTIFICIAL INTELLIGENCE APPROACHES

Dr. Nityashree Nadar
Information Technology, S.I.W.S. N.R. Swamy College of Commerce & Smt. Thirumalai College of Science, Wadala Mumbai – 31, India.
Email: nityashreendar15@gmail.com

Abstract
Emotion recognition using Artificial Intelligence (AI) and Machine Learning (ML) has emerged as a vital area of research, driven by its potential applications in human-computer interaction, healthcare, and beyond. However, developing accurate and robust emotion recognition systems remains challenging due to the complexity of emotions, data quality issues, and the need for real-time processing. This paper addresses these challenges by exploring recent advancements in emotion recognition across various modalities, including facial expressions, speech, and EEG signals. The primary objectives are to review state-of-the-art approaches, identify existing limitations, and propose potential improvements for future research. The methodology involves a comprehensive analysis of recent literature, focusing on models that utilize Deep Learning (DL) techniques such as Convolutional Neural Networks (CNNs) and Graph Convolutional Networks (GCNs). Findings indicate that while significant progress has been made, challenges related to data diversity, multimodal integration, and real-world applicability persist. Ethical considerations, including privacy concerns and bias, are also critical areas of concern. In conclusion, addressing these issues will be essential for advancing emotion recognition technologies and ensuring their safe and effective deployment in real-world applications.
Keywords: Emotion Recognition, Facial Expressions, Speech Emotion Recognition, EEG, Multimodal Integration.

► *Corresponding Author: Dr. Nityashree Nadar*

1. Introduction

An individual's emotional state is often described as a perspective towards the outside world and reality as a whole, which could be negative or positive. Emotions are a collection of mental activities that exhibit as emotions, passions, or feelings [1]. As a general term, "emotions" encompass not just the feelings experienced by individuals but additionally their emotional, mental, and psychological aspects [2]. According to [3], "patients" who would normally exhibit ordinary or possibly exceptional cognitive abilities and reasoning abilities are impacted because of an imbalance in the connection among the limbic network and cerebral cortex, which controls emotions. Nevertheless, their capacity for making choices has faced major challenges [4], demonstrating that an individual's cognitive ability cannot be solely demonstrated by typical cognitive abilities and logic-based skills, but additionally through extensive emotion capabilities. Emotion serves as a significant method of interacting with others, conveying data pertaining to how someone feels and describing intricate psychological processes and psychological intentions through physiological signs like texts [5], changes in intonation-volume [6], [7], face expressions [7], body position [8], and brain activity [9].

Various motor and coding units work together for capturing emotion, allowing for the formation of complicated expression variations on a person’s face [10]. For, example, the Facial-Motion Coding-System (FACS) [11] has the ability to evaluate emotions by employing Deep-Region and Multi-Label-Learning (DRML) [12] structure. It does this by employing feed-forward operations for inducing essential facial structures and by allowing it to generate weights for capturing structural-data related to the human face in accordance with what emotions are collected. Moreover, when it comes to the mechanism of human-knowledge exchange, speech remains the language that individuals use on a regular basis when interacting with one another. Speech, which is one of the most fundamental forms of audio-visual communication, is not just capable of distinguishing between various vocalists but also of efficiently differentiating between different emotions [13]. Research conducted on a global scale emphasizes the investigation of acoustical characteristics of emotive speech, including but not limited to melody, sounds origin, resonating amplitude and bandwidth, and other similar characteristics [14].

Deep-Learning (DL) has emerged as the subject of extensive research over the past decade and offers an extensive number of potentials uses in the field of Speech-Emotion-Recognition (SER) computations [15]. Furthermore, during interactions through SER, a person’s posture frequently conveys details regarding the emotions being spoken [16]. Combining human gaits with Artificial-Intelligence (AI), researchers have developed a number of sensors of motion along with body data collection sensors that can statistically evaluate an individual's external-representation of their psychological state in response to various stimuli [17], [18]. One such system is the smart seat, which is installed in the driver's seat, can detect and respond according to the driver's psychological condition in real-time [19]. Another such example includes a graph known as an Electro-Encephalo-Gram (EEG) which is created by using an accurate electronic device that improves capturing the brain's spontaneously physiological activity from the head [20]. Within the realm of recognizing emotions, it has seen a significant amount of application. The EEG is utilized in studies pertaining to depression [21], consciousness determining detections [22], and therapeutic interventions [23].

According to the theory of music emotions, listening to music can cause listeners to experience a range of feelings [24]. Also, according to scientific research, human bodies release a health-promoting chemical when exposed to soothing musical environments, which in turn reduces mental health issues like stress, depression, anxiousness, and more [25]. The laws governing the musical tastes of individuals have been thoroughly tested through cognitive thinking and information analysis, and an accurate musical signaling value has been developed for assessing the level to which it is aesthetically pleasing [26]. In the recent years, AI has been used to understand the music emotion [27], [28]. This demonstrates how Machine-Learning (ML) can assess how individuals feel in relation to music; future studies can build on this principle to replicate high-quality music which adheres to societal cultural norms and to bring about the connection among technology and emotions.

For capturing both cognitive and physiological mechanisms linked with an individuals present emotional state, this work aims at studying the emotion using various approaches presented in recent years using AI. This work investigates various methods for recognizing emotions by utilizing various kinds that has been gathered while taking into consideration various audio-visual properties. In addition, the survey sheds insight on the limitations of the methodology and datasets that are currently being utilized, which can be taken into consideration in the future for improved emotion recognition. In this manuscript, the structure is organized as follows. Section II provide a comprehensive literature survey, highlighting recent advancements in emotion recognition using AI and ML across various modalities, including facial expressions, speech, and EEG signals. The section also discusses the methodologies and techniques employed in these studies, as well as their reported performances. Section III discusses the findings from the literature survey, synthesizing the results from various studies to identify trends, commonalities, and key insights. This section also compares the performance of different models and approaches across different datasets and tasks. In Section IV, the issues and challenges faced in the field of emotion recognition are discussed. This section addresses the limitations identified in the literature, including data quality, multimodal integration, real-world applicability, and ethical concerns. Potential solutions and areas for future research are also explored. Finally, in Section V, the conclusion of the work is provided, summarizing the key findings, challenges, and the potential future directions for advancing emotion recognition technologies. This section also emphasizes the importance of addressing the identified issues to achieve more robust and ethical AI-driven emotion recognition systems.

2. Literature Survey

This section presents a discussion on the recent work presented for emotion recognition and detection using various approaches. In [29], presented a ML-based approach called as Optimal-Graph-Coupled Semi-Supervised-Learning (OGSSL) for recognizing emotions using Electroencephalogram (EEG) signals. This work utilized the advantages of Adaptive-Graph-Learning (AGL) and enhanced their labelling of emotions using an indication matrix. By using the OGSSL-AGL matrix, the EEG frequency bands along with brain areas were identified. Findings were evaluated using SEED-IV dataset [30], where achieved three average cross-session emotion-recognition tasks accuracies as81.29%, 77.08% and 76.51%. In [31], presented a Bi-directional Long-Short-Term-Memory (Bi-LSTM) approach for speech-based emotion-recognition by utilizing Self-Attention based Weight-Correction (SAWC). The Bi-LSTM approach extracted text and acoustic features from speech words and segments for training the SAWC and reducing the error for emotion-recognition. Findings were evaluated using Interactive-Emotional-Dyadic Motion-Capture (IEMOCAP) dataset [32] where achieved 76.6% accuracy for emotion recognition. In [33], utilized DL approaches, i.e., CNN architecture-based methods, VGG16 and AlexNet for face-recognition and emotion-recognition in humanoid-robots. For this, they created a dataset which was collected from 30 electrical-engineering students, applied preprocessing and labelled the dataset. They collected a total of 18,900 images in which only considered 5,000 images for testing and training. Real-time tests were carried out, where achieved 100% and 85% for face-recognition using VGG16 and AlexNet and for emotion-recognition achieved 73% and 64% respectively. In [34], presented a systematic-review for the studies related to multi-modal emotion-recognition approaches. They conducted study of existing approaches presented from 2016 to 2021, where considered work done on face-emotion-recognition, speech-emotion-recognition and EEG-based emotion-recognition. This work discussed the different ML and DL approaches used and concluded by providing the performance of different classifiers using different datasets.

In [35], presented a DL algorithm for speech-emotion-recognition called as F-emotion which helps in selecting the best features for speech-emotion-recognition. In F-emotion, extracted speech-based features, then evaluated a score called as F-emotion for every feature. This score was used for evaluating the speech-emotion-recognition. For training and testing different types of emotions, presented a DL approach where applied a decision-fusion for attaining total speech-emotion-recognition. Findings were evaluated using EMO-DB and RAVDESS dataset [36], [37], where achieved 88.8% and 82.3% accuracy for recognition. In [38], presented a framework which consisted of three phases, where in first phase CNN and Random-Forest (CNN-RF) were combined for emotion-recognition in students during assessment and online-learning. The evaluation for the CNN-RF approach was done using FER-dataset [39], where achieved 71% accuracy. The second phase included the mapping of emotion to instructor where presented a Plutchik's-Wheel based mapping approach. The third phase included a dashboard where showed the visual performance of student using emotional data. In [40], presented an approach called as Multi-Scale-Feature-Reconstruction Graph-Convolutional-Network (MSFR-GCN) for identifying cognition-tasks and emotions. The MSFR included a feature extraction approach where had two modules, Multi-Scale Sample-Reweighting (MSSR) and Multi-Scale Squeeze-Excitation (MSSE). The role of MSSR was to assign sample-weights on basis of combined frequency and channel data and MSSE role was to assign weights for frequency-bands and channels on basis of statistical data. Further, a feature-pool was utilized where GCN was applied to attain EEG channel-data. Evaluations were conducted using SEED-IV and SEED dataset, where achieved 89.02% and 96.63% accuracy for emotion identification. They also tested on Emotion and Cognition-EEG dataset (ECED), which they collected themselves, where achieved 56.88% accuracy for emotion-recognition and 64.10% accuracy for cognition-task.

In [41], presented an approach for identifying emotions using odors, where the subject was asked to smell and based on expression, the video was considered and evaluated. There were total of four emotions, which were considered for evaluation. Evaluations were conducted using oldfactory-enhanced videos where achieve emotion-recognition accuracies varying from 50.54% to 57.55% for different emotions. In [42], presented a speech-emotion-recognition cross-corpus framework which was built on the concept of local-domain adaptation. A categorized-level discrepancy approach was utilized for evaluation of speech-emotion-recognition cross-corpus. For generalizing the two domains, used a local-adaptive approach. For evaluations used Emo-DB [37] and IEMOCAP, where tested three set of experiments and evaluated their model and compares, where achieved better results for different emotions. In [43], presented a spatial-temporal emotion-based recognition approach, which consisted of a hierarchical-self-attention-based network which extracted global and local emotion-level data and reduces noise at temporal-level. For further feature extraction utilized Squeeze-Excitation model incorporated with Channel-Correlation-Loss for selecting task-based channels. Also, presented a localization approach, called as emotion-localization which localized fragments with better emotions. Evaluations were done using SEED, DEAP [44] and MAHNOB-HCI [45]. Findings showed better outcomes in comparison with existing approaches. In [46], utilized CNN architecture for emotion-recognition for detection of human-emotions. The evaluations were conducted using FER-2013 dataset, where achieved accuracy levels varying from 83% to 95% for six emotions. In [47], presented a emotion-recognition framework called as Personality-Enhanced-Iterative-Refinement Network (PIRNet), which captured person emotions using personality-traits. Then DL approach was used for exploiting sequences and capturing contextual data. Evaluations were done using CM-MOSI [48], CMU-MOSEI [49] and IEMOCAP, where findings showed better outcomes for the PIRNet in comparison with other approaches.

In [50], utilized two DL approaches RoBERTa and Inception ResNet-V2, combined them and presented an approach called as RobinNet. This was done to extract text and audio features using a classifier for which utilized intermediate-fusion. Moreover, Inception ResNetV2 was utilized for identifying the speech using speaker suing spectrogram-augmentation. They evaluated their RobinNet using EMOCAP, where achieved 72.8% accuracy for classification. Also, evaluations were done using Multi-Modal Emotion-Lines Dataset (MELD), IEMOCAP, CMU-MOSEL where RobinNet achieved better outcomes. In [52], proposed a DL approach which was constructed using Gated-Recurrent-Unit (GRU) and MulitHead Attention (MHA) network for speech-emotion-recognition. Evaluations were conducted using Emo-DB and IEMOCAP, where achieved 83.22% and 71.01% for recognition tasks. In [53], studied the automatic human emotion-recognition ML approaches where evaluated the different approaches, compared their performance and the different datasets present for evaluation. In [54], presented a MER2024 dataset which has various speech-based vocabulary emotion-recognition. They also presented a tool for evaluation and testing the emotion-recognition. In [55], presented a multi-modal emotion-recognition system for real-time using EEG signals where they used LSTM for feature extraction. They considered a 14 EEG data collected from 14 patients where they were able to classify emotions at accuracy of 90.25% using facial-data and 87.25% accuracy for EEG data. By combining both facial-data and EEG data achieved 99.3% accuracy. In [56], presented a EMOTIC dataset, which has various images from different dataset. They used two DL approaches, Deep-CNN (DCNN) and VGG19 for evaluating their model. They optimized the VGG19 and DCNN approaches for achieving better results. They achieved Mean-Average-Precision of 42.81% for DCNN and 44.12% for VGG19 respectively. In [57], presented an approach called as EmoBox for speech-emotion multi-corpus recognition which utilized an emotion2vec for avoiding errors during recognition and achieved better results for recognition. In [58], evaluated the speed and accuracy of emotion-recognition for which created a dataset having participants from 19-28 having 6 emotions in mask and without mask. Findings showed that, the assessment showed that for participants without masks achieved 85.7% accuracy.

3. Findings

From the above literature survey, the findings are presented in Table 1, where approach used, dataset and achieved accuracy are discussed. Further, in Table 2, the limitations of the different studies are presented.

Table 1. Findings.

Ref	Approach	Dataset	Accuracy
[29]	OGSSL Model for EEG emotion recognition	SEED-IV	76.51%, 77.08%, 81.29% (cross-session)
[31]	BiLSTM and Self-Attention with SAWC for SER	IEMOCAP	76.6% (weighted average)
[33]	CNN Architectures for Face and Emotion Recognition in Humanoid Robots	Custom Dataset	Face: 85% (AlexNet), 100% (VGG16); Emotion: 64% (AlexNet), 73% (VGG16)
[35]	F-Emotion Algorithm for SER	RAVDESS, EMO-DB	82.3% (RAVDESS), 88.8% (EMO-DB)
[38]	CNN-RF for Emotion Detection in Online Learning	FER-2013	71%
[40]	MSFR-GCN for Emotion Recognition	SEED, SEED-IV, ECED	SEED: 96.63%, SEED-IV: 89.02%, ECED: 56.88%

			(emotion), 64.10% (cognition)
[41]	Olfactory-Enhanced Videos for Emotion Recognition	Custom EEG-EOG Dataset	OVEP: 50.54%, OVLP: 51.49%, TVEP: 40.22%, TVLP: 57.55%
[42]	Local Domain Adaptation for Cross-Corpus SER	IEMOCAP, Emo-DB	89% (IEMOCAP), 76% (Emo-DB)
[43]	Two-Step Spatial-Temporal Emotion Recognition	SEED, DEAP, MAHNOB-HCI	73% (SEED), 78% (DEAP), 81% (MAHNOB-HCI)
[46]	CNN for Emotion Detection in Video Content	FER-2013	83%-95%
[47]	PIRNet for Emotion Recognition in Conversations	IEMOCAP, CMU-MOSI, CMU-MOSEI	93% (IEMOCAP), 92% (CMU-MOSI), 92.5% (CMU-MOSEI)
[50]	RobinNet for Multimodal SER	IEMOCAP, MELD, CMU-MOSEI	72.8% (weighted accuracy)
[52]	Bi-GRU and Multi-Head Attention for SER	IEMOCAP, Emo-DB	71.01% (IEMOCAP), 83.22% (Emo-DB)
[55]	Multimodal Emotion Recognition using EEG and Facial Features	Custom Dataset	99.3% (multimodal), 90.25% (facial), 87.25% (EEG)
[56]	DCNN and VGG19 for Emotion Recognition in Images	EMOTIC	Improved Mean Average Precision (mAP) by 42.81% (DCNN), 44.12% (VGG19)
[57]	EmoBox for Multilingual Multi-Corpus SER	32 Emotion Datasets	Intra-corpus and cross-corpus results varied
[58]	Impact of Mask-Wearing on Emotion Recognition	Custom Dataset	Unmasked: 85.7%, Masked: 73.8%

Table 2. Limitations.

Ref	**Limitations**
[29]	Focused on EEG data, limiting its application to other modalities. The performance in uncontrolled environments is not evaluated.
[31]	Method heavily relies on the accuracy of speech recognition; errors in speech recognition can affect emotion detection accuracy.
[33]	Emotion recognition accuracy in humanoid robots remains relatively low compared to face recognition; limited to a small dataset and specific emotions.
[35]	Limited to speech emotion recognition; may not perform well in real-world scenarios with noisy data or mixed emotions.
[38]	Accuracy is moderate, and the method may not handle complex emotions well; the dataset used (FER-2013) may not represent real-world variability.

[40]	Performance drops significantly when applied to the self-collected dataset (ECED), suggesting limited generalizability to diverse datasets.
[41]	Olfactory-enhanced videos have limited applicability and the approach showed moderate accuracy, which might not be sufficient for practical use.
[42]	The method's effectiveness is significantly dependent on the similarity between the source and target domains; may struggle with highly dissimilar datasets.
[43]	While outperforming state-of-the-art methods, the approach might be computationally intensive and not suitable for real-time applications.
[46]	The approach's accuracy varies significantly, and the study is limited to basic emotions; may not capture more complex emotional states.
[47]	Model performance is dependent on the quality of personality trait data; potential challenges in generalizing to broader contexts.
[50]	Intermediate fusion improves performance but may not be as effective in highly noisy or diverse datasets; limited to specific benchmarks.
[52]	The method, while effective, may not scale well to real-time applications or very large datasets; accuracy is limited to the datasets used.
[55]	The approach is dependent on high-quality EEG and facial data; may not perform well in real-world conditions where data quality is variable.
[56]	Performance improvements are dataset-specific; may not generalize well to unseen data or different contexts.
[57]	The toolkit's performance varies across datasets and languages; not all results are reported in terms of specific accuracy, making comparisons challenging.
[58]	Mask-wearing significantly reduces emotion recognition accuracy; limited to specific emotions and controlled conditions, which may not reflect real-world scenarios.

In the next section, the current issues and challenges in emotion-recognition are discussed.

4. Issues and Challenges

The issues and challenges of the work are as follows

1. Data Quality and Diversity:

- **Limited Datasets:** Many emotion recognition models are trained on limited datasets that may not fully represent the diversity of human emotions across different cultures, contexts, and populations.
- **Imbalanced Data:** Datasets often contain imbalances, with some emotions being underrepresented. This can lead to biased models that perform poorly on less common emotions.
- Data Labeling: Emotions are inherently subjective, making it difficult to label data consistently. This subjectivity can lead to noisy labels, reducing model accuracy.

2. Multimodal Integration:

- **Complexity of Fusion:** Combining data from multiple modalities (e.g., facial expressions, speech, EEG) poses challenges in terms of data synchronization, feature extraction, and model fusion.
- **Cross-Modality Discrepancies:** Different modalities might provide conflicting information about the emotional state, making it difficult for models to make accurate predictions.

3. Real-World Applicability:

- **Controlled vs. Uncontrolled Environments:** Many models perform well in controlled environments but struggle in real-world settings with variable lighting, background noise, and other distractions.
- **Generalization to New Contexts:** Models trained on specific datasets may not generalize well to new or unseen contexts, limiting their practical use in diverse real-world applications.

4. Computational Complexity:

- **Resource-Intensive Models:** DL models, especially those involving complex architectures like CNNs or GCNs, can be computationally intensive, requiring significant hardware resources for real-time applications.
- **Latency Issues:** Real-time emotion recognition is challenging due to the need for quick processing and low latency, especially in applications like human-robot interaction.

5. Ethical and Privacy Concerns:

- **Privacy Invasion:** Emotion recognition systems, particularly those using facial recognition or EEG, raise concerns about privacy, especially when used without explicit consent.
- **Bias and Fairness:** AI models can inherit biases present in the training data, leading to unfair treatment of certain groups based on race, gender, or cultural background.

6. Interpretation and Explainability:

- **Black-Box Models:** Many DL models are often viewed as "black boxes," making it difficult to understand how they arrive at a particular decision, which can be problematic for trust and transparency.
- **Interpretation of Results:** The interpretation of detected emotions can be ambiguous, especially in cases where multiple emotions overlap or where the intensity of emotions varies.

7. Emotion Complexity:

- **Subtle Emotions:** Recognizing subtle or mixed emotions remains a significant challenge, as most models are designed to detect basic emotions like happiness, sadness, or anger.
- **Temporal Dynamics:** Emotions are dynamic and can change rapidly over time. Capturing these temporal dynamics accurately is crucial but challenging.

8. Cross-Cultural Variability:

- **Cultural Differences:** Emotional expressions and perceptions vary widely across cultures, making it difficult to develop universal models that work equally well in all cultural contexts.
- **Localization:** Adapting emotion recognition models to different cultural contexts requires localization, which is resource-intensive and may not always be feasible.

9. Integration with Other Systems:

- **Interoperability:** Emotion-recognition systems need to be integrated with other AI systems (e.g., conversational agents, autonomous vehicles), posing challenges in terms of interoperability and system architecture.
- **User Acceptance:** The effectiveness of emotion-recognition systems also depends on user acceptance, which can be hindered by concerns over privacy, accuracy, and the system's perceived intrusiveness.

10. Ethical Implications and Societal Impact:

- **Misuse Potential:** There is a risk of misuse of emotion-recognition technologies, such as in surveillance or manipulative advertising, raising ethical concerns.
- **Impact on Mental Health:** Prolonged use of emotion recognition systems, particularly in sensitive contexts, could have unintended consequences on users' mental health and well-being.

5. Conclusion

The exploration of emotion-recognition and detection using AI and ML has shown remarkable progress in recent years, with significant advancements across various modalities, including facial expressions, speech, EEG signals, and multimodal approaches. The integration of sophisticated models, such as CNNs, RNNs, and GCNs, alongside innovative techniques like self-attention mechanisms, domain adaptation, and TL, has enhanced the accuracy and robustness of emotion recognition systems. However, despite these advancements, several challenges remain. The quality and diversity of datasets continue to pose limitations, with issues related to imbalanced data, subjective labeling, and the lack of representation of diverse populations. Moreover, the complexity of fusing multimodal data, the difficulty of deploying models in real-world, uncontrolled environments, and the computational demands of deep learning architectures highlight the need for further research and development. Ethical concerns, particularly around privacy, bias, and the potential misuse of emotion recognition technologies, also underscore the importance of developing transparent, fair, and responsible AI systems. Future research should focus on addressing these challenges by developing more generalized and interpretable models, enhancing the diversity and quality of datasets, and creating ethical frameworks for the deployment of emotion recognition systems. By overcoming these issues, the field can move closer to realizing the full potential of emotion recognition technologies in various applications, from healthcare and education to human-computer interaction and beyond.

References

1. G. E. Marcus, “Evaluating the status of theories of emotion in political science and psychology,” Frontiers in Political Science, vol. 4, Feb. 2023, doi: 10.3389/fpos.2022.1080884.
2. Paolo Cianconi, Batul Hanife, F. Grillo, S. Betro, B. Jaya, and L. Janiri, “Eco-emotions and Psychoterratic Syndromes: Reshaping Mental Health Assessment Under Climate Change,” Yale Journal of Biology and Medicine, vol. 96, no. 2, pp. 211–226, Jun. 2023, doi: 10.59249/earx2427.
3. Z. Qi, Y. An, M. Zhang, H.-J. Li, and J. Lu, “Altered Cerebro-Cerebellar Limbic Network in AD Spectrum: A Resting-State fMRI Study,” Frontiers in Neural Circuits, vol. 13, Nov. 2019, doi: 10.3389/fncir.2019.00072.
4. X. Cui, Y. Wu, J. Wu, Z. You, J. Xiahou, and M. Ouyang, “A review: Music-emotion recognition and analysis based on EEG signals,” Frontiers in Neuroinformatics, vol. 16, Oct. 2022, doi: 10.3389/fninf.2022.997282.
5. Kristína Machová, M. Szabóová, Ján Paralič, and Ján Mičko, “Detection of emotion by text analysis using machine learning,” Frontiers in Psychology, vol. 14, Sep. 2023, doi: 10.3389/fpsyg.2023.1190326.
6. K. Yang et al., “Behavioral and Physiological Signals-Based Deep Multimodal Approach for Mobile Emotion Recognition,” IEEE Transactions on Affective Computing, pp. 1–1, 2021, doi: 10.1109/taffc.2021.3100868.
7. Y. Cai, X. Li, and J. Li, “Emotion Recognition Using Different Sensors, Emotion Models, Methods and Datasets: A Comprehensive Review,” Sensors, vol. 23, no. 5, p. 2455, Feb. 2023, doi: 10.3390/s23052455.
8. H. Riemer, J. V. Joseph, A. Y. Lee, and R. Riemer, “Emotion and motion: Toward emotion recognition based on standing and walking,” PLOS ONE, vol. 18, no. 9, pp. e0290564–e0290564, Sep. 2023, doi: 10.1371/journal.pone.0290564.
9. M. -C. Lo, R. Younk and A. S. Widge, "Paired Electrical Pulse Trains for Controlling Connectivity in Emotion-Related Brain Circuitry," in IEEE Transactions on Neural Systems and

Rehabilitation Engineering, vol. 28, no. 12, pp. 2721-2730, Dec. 2020, doi: 10.1109/TNSRE.2020.3030714.
10. Jin Pyo Lee et al., “Encoding of multi-modal emotional information via personalized skin-integrated wireless facial interface,” Nature Communications, vol. 15, no. 1, Jan. 2024, doi: 10.1038/s41467-023-44673-2.
11. T. A. Höfling and G. W. Alpers, “Automatic facial coding predicts self-report of emotion, advertisement and brand effects elicited by video commercials,” Frontiers in Neuroscience, vol. 17, May 2023, doi: 10.3389/fnins.2023.1125983.
12. K. Zhao, W. -S. Chu and H. Zhang, "Deep Region and Multi-label Learning for Facial Action Unit Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 3391-3399, doi: 10.1109/CVPR.2016.369.
13. Juan-Miguel López-Gil and N. Garay-Vitoria, “Assessing the effectiveness of ensembles in Speech Emotion Recognition: Performance analysis under challenging scenarios,” Expert Systems with Applications, vol. 243, pp. 122905–122905, Jun. 2024, doi: 10.1016/j.eswa.2023.122905.
14. K. Bhangale and M. Kothandaraman, “Speech Emotion Recognition Based on Multiple Acoustic Features and Deep Convolutional Neural Network,” Electronics, vol. 12, no. 4, p. 839, Feb. 2023, doi: 10.3390/electronics12040839.
15. D. Li, Y. Zhou, Z. Wang, and D. Gao, “Exploiting the potentialities of features for speech emotion recognition,” Information Sciences, vol. 548, pp. 328–343, Feb. 2021, doi: 10.1016/j.ins.2020.09.047.
16. J. Keck, A. Zabicki, J. Bachmann, J. Munzert, and B. Krüger, “Decoding spatiotemporal features of emotional body language in social interactions,” Scientific Reports, vol. 12, no. 1, p. 15088, Sep. 2022, doi: 10.1038/s41598-022-19267-5.
17. R. Abdulghafor, A. Abdelmohsen, S. Turaev, M. A. H. Ali, and S. Wani, “An Analysis of Body Language of Patients Using Artificial Intelligence,” Healthcare, vol. 10, no. 12, p. 2504, Dec. 2022, doi: 10.3390/healthcare10122504.
18. E. J. Harris, I-Hung. Khoo, and E. Demircan, “A Survey of Human Gait-Based Artificial Intelligence Applications,” Frontiers in Robotics and AI, vol. 8, Jan. 2022, doi: 10.3389/frobt.2021.749274.
19. H. Xiao et al., “On-Road Driver Emotion Recognition Using Facial Expression,” Applied Sciences, vol. 12, no. 2, p. 807, Jan. 2022, doi: 10.3390/app12020807.
20. A. Chaddad, Y. Wu, Reem Kateb, and A. Bouridane, “Electroencephalography Signal Processing: A Comprehensive Review and Analysis of Methods and Techniques,” Sensors, vol. 23, no. 14, pp. 6434–6434, Jul. 2023, doi: 10.3390/s23146434.
21. H. Chang, Y. Zong, W. Zheng, C. Tang, J. Zhu, and X. Li, “Depression Assessment Method: An EEG Emotion Recognition Framework Based on Spatiotemporal Neural Network,” Frontiers in Psychiatry, vol. 12, p. 837149, Mar. 2022, doi: 10.3389/fpsyt.2021.837149.
22. A. Al-Nafjan, M. Hosny, Y. Al-Ohali, and A. Al-Wabil, “Review and Classification of Emotion Recognition Based on EEG Brain-Computer Interface System Research: A Systematic Review,” Applied Sciences, vol. 7, no. 12, p. 1239, Dec. 2017, doi: 10.3390/app7121239.
23. H. Zhou, W. Liang, H. Liu, L. Wang, K. H. Ryu, and K. W. Nam, “EEG Emotion Recognition Applied to the Effect Analysis of Music on Emotion Changes in Psychological Healthcare,” International Journal of Environmental Research and Public Health, vol. 20, no. 1, p. 378, Jan. 2023, doi: 10.3390/ijerph20010378.

24. H. Strauss et al., “The Emotion-to-Music Mapping Atlas (EMMA): A systematically organized online database of emotionally evocative music excerpts,” Behavior Research Methods, Jan. 2024, doi: 10.3758/s13428-024-02336-0.
25. C. Eseadi and M. O. Ngwu, “Significance of music therapy in treating depression and anxiety disorders among people with cancer,” World Journal of Clinical Oncology, vol. 14, no. 2, pp. 69–80, Feb. 2023, doi: 10.5306/wjco.v14.i2.69.
26. O. Herdson, T. Eerola, and A.-H. Javadi, “Analysis and Classification of Music-Induced States of Sadness,” Emotion Review, p. 175407392211404, Nov. 2022, doi: 10.1177/17540739221140472.
27. X. Han, F. Chen, and J. Ban, “Music Emotion Recognition Based on a Neural Network with an Inception-GRU Residual Structure,” Electronics, vol. 12, no. 4, p. 978, Feb. 2023, doi: 10.3390/electronics12040978.
28. J. Lian, “An artificial intelligence-based classifier for musical emotion expression in media education,” PeerJ Computer Science, vol. 9, p. e1472, Jul. 2023, doi: 10.7717/peerj-cs.1472.
29. Y. Peng, F. Jin, W. Kong, F. Nie, B. -L. Lu and A. Cichocki, "OGSSL: A Semi-Supervised Classification Model Coupled With Optimal Graph Learning for EEG Emotion Recognition," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 30, pp. 1288-1297, 2022, doi: 10.1109/TNSRE.2022.3175464.
30. “SEED Dataset,” Sjtu.edu.cn, 2019. https://bcmi.sjtu.edu.cn/home (accessed Aug. 08, 2024).
31. J. Santoso, T. Yamada, K. Ishizuka, T. Hashimoto and S. Makino, "Speech Emotion Recognition Based on Self-Attention Weight Correction for Acoustic and Text Features," in IEEE Access, vol. 10, pp. 115732-115743, 2022, doi: 10.1109/ACCESS.2022.3219094.
32. “IEMOCAP- Home,” Usc.edu, 2022. https://sail.usc.edu/iemocap/
33. S. Dwijayanti, M. Iqbal and B. Y. Suprapto, "Real-Time Implementation of Face Recognition and Emotion Recognition in a Humanoid Robot Using a Convolutional Neural Network," in IEEE Access, vol. 10, pp. 89876-89886, 2022, doi: 10.1109/ACCESS.2022.3200762.
34. M. M et al., “Automated Emotion Recognition: Current Trends and Future Perspectives,” Computer Methods and Programs in Biomedicine, vol. 215, p. 106646, Jan. 2022, doi: 10.1016/j.cmpb.2022.106646.
35. L. -M. Zhang, G. W. Ng, Y. -B. Leau and H. Yan, "A Parallel-Model Speech Emotion Recognition Network Based on Feature Clustering," in IEEE Access, vol. 11, pp. 71224-71234, 2023, doi: 10.1109/ACCESS.2023.3294274.
36. S. R. Livingstone and F. A. Russo, "The Ryerson audio-visual database of emotional speech and song (RAVDESS): A dynamic multimodal set of facial and vocal expressions in North American English", PLoS ONE, vol. 13, no. 5, May 2018.
37. F. Burkhardt, A. Paeschke, M. Rolfes, W. F. Sendlmeier and B. Weiss, "A database of German emotional speech", Proc. Interspeech, vol. 5, pp. 1517-1520, Sep. 2005.
38. A. Sassi, W. Jaafar, S. Cherif, J. B. Abderrazak and H. Yanikomeroglu, "Video Traffic Analysis for Real-Time Emotion Recognition and Visualization in Online Learning," in IEEE Access, vol. 11, pp. 99376-99386, 2023, doi: 10.1109/ACCESS.2023.3313973.
39. I. J. Goodfellow et al., "Challenges in representation learning: A report on three machine learning contests", Neural Information Processing, pp. 117-124, 2013.
40. D. Pan et al., "MSFR-GCN: A Multi-Scale Feature Reconstruction Graph Convolutional Network for EEG Emotion and Cognition Recognition," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 31, pp. 3245-3254, 2023, doi: 10.1109/TNSRE.2023.3304660.

41. M. Wu, W. Teng, C. Fan, S. Pei, P. Li and Z. Lv, "An Investigation of Olfactory-Enhanced Video on EEG-Based Emotion Recognition," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 31, pp. 1602-1613, 2023, doi: 10.1109/TNSRE.2023.3253866.
42. Z. Huijuan, Y. Ning and W. Ruchuan, "Improved Cross-Corpus Speech Emotion Recognition Using Deep Local Domain Adaptation," in Chinese Journal of Electronics, vol. 32, no. 3, pp. 640-646, May 2023, doi: 10.23919/cje.2021.00.196.
43. Y. Zhang, H. Liu, D. Zhang, X. Chen, T. Qin and Q. Zheng, "EEG-Based Emotion Recognition With Emotion Localization via Hierarchical Self-Attention," in IEEE Transactions on Affective Computing, vol. 14, no. 3, pp. 2458-2469, 1 July-Sept. 2023, doi: 10.1109/TAFFC.2022.3145623.
44. Sander Koelstra, "DEAP: A Dataset for Emotion Analysis using Physiological and Audiovisual Signals," Qmul.ac.uk, 2011. https://www.eecs.qmul.ac.uk/mmv/datasets/deap/ (accessed Aug 08, 2024).
45. Mahnob-db.eu, 2024. https://mahnob-db.eu/
46. J. A. Ballesteros, G. M. Ramírez, F. Moreira, A. Solano, and C. A. Pelaez, "Facial emotion recognition through artificial intelligence," Frontiers in computer science, vol. 6, Jan. 2024, doi: 10.3389/fcomp.2024.1359471.
47. Z. Lian, B. Liu and J. Tao, "PIRNet: Personality-Enhanced Iterative Refinement Network for Emotion Recognition in Conversation," in IEEE Transactions on Neural Networks and Learning Systems, vol. 35, no. 2, pp. 2863-2874, Feb. 2024, doi: 10.1109/TNNLS.2022.3192469.
48. "CMU-MOSI Dataset | MultiComp," MultiComp | MultiComp Lab's mission is to build the algorithms and computational foundation to understand the interdependence between human verbal, visual, and vocal behaviors expressed during social communicative interactions., Jan. 13, 2018. http://multicomp.cs.cmu.edu/resources/cmu-mosi-dataset/ (accessed Aug. 08, 2024).
49. [49] "CMU-MOSEI Dataset | MultiComp," MultiComp | MultiComp Lab's mission is to build the algorithms and computational foundation to understand the interdependence between human verbal, visual, and vocal behaviors expressed during social communicative interactions., Feb. 10, 2018. http://multicomp.cs.cmu.edu/resources/cmu-mosei-dataset/
50. Y. Khurana, S. Gupta, R. Sathyaraj and S. P. Raja, "RobinNet: A Multimodal Speech Emotion Recognition System With Speaker Recognition for Social Interactions," in IEEE Transactions on Computational Social Systems, vol. 11, no. 1, pp. 478-487, Feb. 2024, doi: 10.1109/TCSS.2022.3228649.
51. "MELD," Github.io, 2018. https://affective-meld.github.io/ (accessed Aug. 08, 2024).
52. C. Xu, Y. Liu, W. Song, Z. Liang, and X. Chen, "A New Network Structure for Speech Emotion Recognition Research," Sensors, vol. 24, no. 5, pp. 1429–1429, Feb. 2024, doi: 10.3390/s24051429.
53. Eman, S. Mohsen, E. H. Hussein, and A. Sadek, "Machine learning for human emotion recognition: a comprehensive review," Neural computing & applications, Feb. 2024, doi: 10.1007/s00521-024-09426-2.
54. Z. Lian et al., "MER 2024: Semi-Supervised Learning, Noise Robustness, and Open-Vocabulary Multimodal Emotion Recognition," arXiv.org, May 23, 2024. https://arxiv.org/abs/2404.17113 (accessed Aug 08, 2024).
55. A.M. Mutawa and A. Hassouneh, "Multimodal Real-Time patient emotion recognition system using facial expressions and brain EEG signals based on Machine learning and Log-Sync methods," Biomedical signal processing and control, vol. 91, pp. 105942–105942, May 2024, doi: 10.1016/j.bspc.2023.105942.

56. Fatiha Limami, Boutaina Hdioud, and Rachid, "Contextual emotion detection in images using deep learning," Frontiers in Artificial Intelligence, vol. 7, Jun. 2024, doi: 10.3389/frai.2024.1386753.
57. Z. Ma et al., "EmoBox: Multilingual Multi-corpus Speech Emotion Recognition Toolkit and Benchmark," arXiv (Cornell University), Jun. 2024, doi: 10.48550/arxiv.2406.07162.
58. Arben Hysenaj et al., "Accuracy and speed of emotion recognition with face masks," Europe's Journal of Psychology, vol. 20, no. 1, pp. 16–24, Feb. 2024, doi: 10.5964/ejop.11789.