



OPEN ACCESS

This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

¹Department of Electronics and Communication Engineering, St.Xavier's Catholic College of Engineering, Kanyakumari, Tamil Nadu, India

²Department of Mechanical Engineering, Vimal Jyothi Engineering College, Kannur, Kerala, India

³Department of Electronics and Communication Engineering, Saveetha Engineering College, Thandalam, Chennai, Tamil Nadu, India

⁴Directorate of Research & Innovation, CMR University, Bengaluru, Karnataka, India

Correspondence to:
Thomas Mary Little Flower,
mlittleflower@gmail.com

Additional material is published online only. To view please visit the journal online.

Cite this as: Flower TML, Singh SCE, Jaya T and Devadhas GG. EEG-Based Emotion Recognition: A Systematic Review of Traditional and Deep Learning Methods. Premier Journal of Science 2025;15:100180

DOI: <https://doi.org/10.70389/PJS.100180>

Peer Review

Received: 16 August 2025

Last revised: 17 November 2025

Accepted: 23 November 2025

Version accepted: 5

Published: 7 January 2026

EEG-Based Emotion Recognition: A Systematic Review of Traditional and Deep Learning Methods

Thomas Mary Little Flower¹, Sreedharan Christopher Ezhil Singh², Thirasama Jaya³ and George Glan Devadhas⁴

ABSTRACT

OBJECTIVE

This systematic review provides a synthesis of the existing data concerning the Electroencephalogram (EEG)-based emotion recognition and assesses the development of the old machine learning models to the current deep learning models. The purpose of the review is the comparison of their performance and the identification of trends in the approaches to the methodology and the evaluation of the strength and the reproducibility of the discipline.

METHODS

The review was done based on the Preferred Reporting Items of Systematic Reviews and Meta-Analyses (PRISMA 2020) guidelines. Five electronic databases (IEEE Xplore, Scopus, PubMed, ScienceDirect, and Springer-Link) that have been published not earlier than January 2012 were searched systematically. Due to the removal of duplicates and two rounds of screening against pre-defined inclusion criteria, 50 studies were incorporated to be final synthesized.

FINDINGS

It has been demonstrated that there is a definite trend towards end-to-end deep learning models, especially Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and architectures combining both. These models, particularly when using maps of topographic features and maps of functional connectivity, have shown a higher level of performance, and their classification accuracies of 90 percent or higher in benchmark data such as DEAP and SEED in subject-dependent contexts. But, there is a significant decline in the performance in cross-subject validation, which is an outstanding generalization issue. It also becomes evident during the synthesis that validation protocols, data preprocessing and reporting standards exhibit high heterogeneity, thus making it difficult to directly compare them and jeopardizing reproducibility.

CONCLUSION

Deep learning approaches are an important development in emotion recognition of EEG, but the area is plagued by lack of uniformity and focus on real-world applicability. The next step in work is to focus on the creation of standardized evaluation metrics, explicable AI methods, and effective, cross-subject models to enable the movement of laboratory studies to the reliable, deployable systems.

Keywords: Tunable Q wavelet transform, Topographic EEG feature maps, Convolutional fuzzy neural network, EEG graph neural networks, Valence–arousal classification

Introduction

Human emotional states are accurately recognized, which is a fundamental to Human-Computer Interaction (HCI),

brain-computer interfaces (BCIs) and affective computing. Emotions as a complicated psychological and physiological phenomenon have a significant impact on cognition, decision-making, and behavior. Although emotion recognition can be done through different modalities such as facial expression and speech, they are either consciously suppressed or even culturally influenced. An alternative approach, which is more direct with respect to the inner affective states, is electroencephalography (EEG), which offers a non-invasive, high-temporal-resolution window of the electrical activity of the brain.

Theoretical models play a vital role in framing emotional classification. Two commonly used paradigms are the discrete emotion model and the dimensional emotion model. The discrete model, based on the work of Ekman, categorizes emotions into basic types such as happiness, sadness, anger, fear, surprise, and disgust. In contrast, the dimensional model represents emotions along continuous axes, typically valence (positive to negative) and arousal (calm to excite). The dimensional model is particularly well suited to EEG studies, as it aligns with the continuous and dynamic nature of brain activity. EEG signals are characterized by their non-linear and non-stationary nature, making them susceptible to various artifacts and noise, such as those arising from muscle movements, eye blinks, and environmental interferences. These challenges necessitate robust pre-processing techniques to ensure the reliability of the extracted features. Common pre-processing steps include filtering to remove noise, artifact rejection methods, and normalization procedures to standardize the data across different sessions and subjects.

Feature extraction is a critical step in EEG-based emotion recognition, aiming to distill meaningful information from raw EEG signals. Traditional methods involve analyzing the signals in time, frequency, and time–frequency domains. Techniques such as empirical mode decomposition (EMD), wavelet transforms, and Hilbert–Huang transforms have been widely employed to capture the intricate dynamics of EEG signals. These methods decompose the signals into components that reflect various frequency bands associated with different cognitive and emotional states. Advanced signal processing methods, including tunable Q wavelet transform (TQWT) and multivariate synchrosqueezing transform (MSST), offer effective decomposition of EEG signals across frequency bands while preserving temporal information. These techniques provide rich feature sets that enhance classification accuracy by capturing both the spectral and temporal characteristics of the EEG signals.

Feature representation plays a crucial role in enhancing the accuracy of emotion classification.

Ethical approval: The six EEG-based datasets, namely, DEAP, SEED, DREAMER, AMIGOS, MAHNOB-HCI, and ASCERTAIN are publicly accessible and commonly used by researchers to extract features and then classify emotions

Consent: N/a

Funding: No industry funding

Conflicts of interest: N/a

Author contribution:

Thomas Mary Little Flower, Sreedharan Christopher Ezhil Singh, Thirasama Jaya and George Glan Devadhas – Conceptualization, Writing – original draft, review and editing

Guarantor: Thomas Mary Little Flower

Provenance and peer-review: Unsolicited and externally peer-reviewed

Data availability statement: N/a

Several studies have proposed the use of topographic and holographic feature maps derived from EEG signals. These maps encode spatial information by mapping electrode positions onto a two-dimensional grid, thereby preserving the geometric layout of the brain's surface. Additionally, connectivity-based features, which represent functional interactions between brain regions, have gained traction. Measures such as Pearson's correlation coefficient, phase-locking value (PLV), and transfer entropy (TE) have been utilized to construct connectivity matrices, which serve as inputs to deep learning models. These representations capture the dynamic relationships between different brain regions, providing insights into the neural mechanisms underlying emotional processing.

Emotion recognition based on the EEG has developed fast within the last ten years. The first methods were based on a fully developed pipeline: signal pre-processing, time, frequency, time-frequency feature extraction, and classification with the help of classical machine learning models: Support Vector Machines (SVMs) and Random Forests. In more recent times, deep learning architectures, such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), transformers, and others, have appeared, which can automatically extract hierarchical features of raw or minimally processed EEG data. These models have demonstrated great achievement but come with fresh problems in the terms of cost of computation, interpretability, and reproducibility.

Although the current literature provides a great amount of separate studies, it is necessary to conduct a synthesis of the findings in a comprehensive and well-defined study with high quality methodology to sum up all the research results, to evaluate the methodological level of the research critically and to outline the comparative effectiveness of these developing paradigms. Other reviews in the past have typically been narrative in nature, and have not met the systematicity one needs to reduce bias and then give a conclusive report of the evidence picture.

To fill this gap, we have performed a literature review on EEG-based emotion recognition. The main research question in this review is as follows and would be organized into the main essential parts of a systematic review:

- Population: EEG responses to identify emotions.
- Intervention: Deep Learning models (e.g. CNN, LSTM, Transformers).
- Comparison: Conventional Machine Learning models (e.g., SVM, k-NN, Random Forest).
- Outputs: Primary: Accuracy of classification, F1-score. Secondary: Cross-subject, validation protocol transparency.

To achieve the following objectives:

- Methods Systematic search, sifting and synthesizing of pertinent literature that employs traditional and deep learning techniques to recognize emotions using EEG.

- Compare the reported methodological performance (e.g. accuracy, F1-score) of these methods quantitatively on benchmark datasets.
- Critically assess model generalizability and strength, by comparing the model performance on subject-dependent and cross-subject validation environments.
- Determine the methodological transparency and risk of bias used in the literature contained in the studies, paying attention to data split reporting, code availability, and other reproducibility measures.
- Detect existing gaps in research and, following the synthesized evidence, give practical recommendations to the future work.

Literature Survey

EEG-based emotional state understanding not only improves brain-computer interface (BCI) systems but also makes a substantial contribution to adaptive human-computer interaction, individualized education, and mental health diagnoses. EEG-based emotion recognition has advanced significantly over the last ten years, moving from manual feature extraction and classical classifiers to advanced deep learning and hybrid architectures that successfully capture the intricate, non-linear dynamics of EEG signals. A variety of feature extraction approaches An in 2023, machine learning models, hybrid frameworks, and benchmark datasets, as highlighted in a literature review that summarizes significant advancements and methodologies have shaped the present status of EEG-based emotion identification research.

Mert and Akan¹ introduced the Multivariate Synchrosqueezing Transform (MSST) to enhance the time-frequency representation of EEG signals. This method provided compact, high-resolution representations that improved the discrimination of emotional states. Feature dimensionality was further reduced using Independent Component Analysis (ICA) and Non-negative Matrix Factorization (NMF), enabling efficient processing of EEG data. Similarly, Subasi in 2021 proposed a modular pipeline incorporating Multi-Scale Principal Component Analysis (MSPCA) for denoising, Tunable Q Wavelet Transform (TQWT) for signal decomposition, and statistical feature extraction, achieving over 93% accuracy on the DEAP dataset with Rotation Forest ensembles.

Moving beyond single-channel analysis, Moon² adopted a brain-wide functional approach by constructing connectivity matrices using Pearson Correlation Coefficient (PCC), Phase-Locking Value (PLV), and Transfer Entropy (TE). This approach captured inter-channel synchrony and enhanced feature representation, allowing the model to leverage functional brain connectivity for emotion classification.

The shift toward deep learning has significantly enhanced the performance of EEG-based emotion recognition systems. CNNs have become a cornerstone due to their proficiency in extracting spatial and spectral patterns from EEG data. Topic and Russo³ utilized EEG-derived Topographic (TOPO-FM) and Holographic

(HOLO-FM) Feature Maps as 2D CNN inputs, achieving state-of-the-art accuracy across datasets such as DEAP, SEED, DREAMER, and AMIGOS. These 2D maps preserved the geometric relationships among EEG channels, improving spatial coherence in feature learning.

Liu⁴ combined CNNs for automated feature extraction with Support Vector Machines (SVMs) for classification. This hybrid approach yielded superior performance in valence-arousal classification tasks and showed better generalization in subject-independent settings. Ensemble techniques, such as the Rotation Forest proposed by Subasi et al., further improved generalization by integrating diverse base classifiers, including k-NN, SVM, and Artificial Neural Networks (ANNs). Such ensembles outperformed individual classifiers, particularly in cross-subject Kuang⁵ evaluations. Boosting and bagging strategies, when used in conjunction with dimensionality reduction techniques like Principal Component Analysis (PCA) and ICA, have also proven effective for managing the high dimensionality of EEG data. These ensemble methods offer scalable, robust performance and are particularly beneficial in real-world settings where data variability is high.

Azar et al.⁶ proposed a Modified Convolutional Fuzzy Neural Network (MCFNN), which integrated the spatial structure of CNNs with fuzzy logic to better handle the uncertainty inherent in emotional EEG signals. Differential Entropy (DE), a robust frequency-domain feature, was extracted from the DEAP dataset. The MCFNN outperformed standard CNNs by achieving higher classification accuracy and better generalization across subjects. Khan⁷ developed a traditional EEG-based emotion recognition system using statistical moments (mean, standard deviation, skewness, kurtosis) and frequency-domain features like Power Spectral Density (PSD) and band power. Using DEAP and DREAMER datasets, they implemented SVM, k-NN, and Random Forest classifiers. SVMs demonstrated superior performance, reaching over 85% accuracy in valence-arousal classification, confirming that traditional machine learning remains competitive when paired with strong feature engineering.

Rahman⁸ performed a comparative analysis of machine learning and deep learning models. Utilizing time-frequency features such as Discrete Wavelet Transform (DWT) and Short-Time Fourier Transform (STFT), they benchmarked traditional models (SVM, Decision Trees) against CNN and Long Short-Term Memory (LSTM) networks. The CNN-LSTM hybrid model outperformed others by leveraging both spatial and temporal aspects of EEG data. Chen⁹ employed Recurrent Neural Networks (RNNs) enhanced with attention mechanisms to classify emotional states from EEG signals. Using DE features from the DEAP dataset, the attention-enhanced RNN dynamically prioritized relevant time segments, significantly improving classification performance. The CNN-LSTM-Attention hybrid model achieved an impressive accuracy of 94%, underscoring the efficacy of attention mechanisms in modeling temporal EEG dynamics.

Wang¹⁰ developed a hybrid model integrating CNN and LSTM layers. Using STFT-based time-frequency features extracted from the SEED dataset, CNNs captured spatial dependencies across EEG channels, while LSTMs modeled temporal sequences. Their model achieved a classification accuracy of 91%, further validating the complementary strengths of spatial and temporal modeling.

Liu¹¹ integrated firefly optimization algorithms with CNN-GRU networks to improve hyperparameter tuning and feature subset selection. Using DE and wavelet-based features from the DEAP dataset, the firefly algorithm optimized network parameters and improved convergence speed, resulting in over 92% classification accuracy. This metaheuristic approach demonstrated the value of intelligent optimization in enhancing deep learning models.

Singh and Sharma¹² introduced a multi-level feature fusion framework that incorporated time-domain, frequency-domain, and nonlinear entropy features. A Gradient Boosting Machine (GBM) was used for classification on the SEED dataset. The model achieved strong performance in multi-class emotion classification, highlighting the benefits of combining diverse feature types.

Li¹³ provided an extensive review of EEG-based emotion recognition, discussing the efficacy of various feature extraction techniques, including PSD, DE, wavelet coefficients, entropy, and fractal dimension. They compared classifiers such as SVM, k-NN, CNN, and LSTM, and emphasized ongoing challenges, including subject dependency, EEG noise, and limited generalization. Their review identified potential in deep learning models, particularly those capable of automatic feature learning and spatiotemporal modeling.

Patel and Chauhan¹⁴ conducted a systematic review focusing on datasets, feature selection methods, and classifier performance. They noted the dominance of frequency-domain features and emphasized the importance of dimensionality reduction techniques like PCA and mutual information for improving model efficiency. SVM and ensemble classifiers were found to be consistently reliable, while deep learning models such as CNNs and hybrid architectures demonstrated increasing popularity due to their scalability and automation. Alarcão and Fonseca¹⁵ provided a tutorial review that classified EEG features into statistical, spectral, and chaotic categories. Their discussion on classifiers ranged from simple linear models to complex deep networks. They also highlighted critical pre-processing steps and the need for standardization across datasets and evaluation protocols.

Taken together, these works indicate a clear trajectory toward more integrated, flexible, and intelligent systems for EEG-based emotion recognition. From handcrafted features and classical classifiers to hybrid deep learning frameworks optimized with bio-inspired algorithms, the field has matured significantly. The use of connectivity matrices, attention mechanisms, and multi-level feature fusion strategies reflects an

increasing understanding of the neural basis of emotion and the complexities of EEG data.

Methods and PRISMA Workflow

To provide transparency, reproducibility and methodological rigor, this systematic review was done in line with the Preferred Reporting Items of a Systematic Review and Meta-Analysis (PRISMA 2020). A systematic literature search, the screening of eligibility, inclusion/exclusion filtering, and data extraction were included in the workflow as outlined in Figure 1 (PRISMA flow diagram).

Databases and Time Frame

There was a thorough literature search conducted into five major academic databases namely IEEE Xplore, Scopus, PubMed, ScienceDirect and SpringerLink which were chosen to help identify engineering- and biomedical-oriented literature on EEG-based emotion recognition. It was search period of January 2012 to August 2025, which coincided with the release of benchmark EEG datasets (e.g. DEAP, SEED) and the fast development of deep learning architectures.

Search Databases and Dates

The literature search was conducted across five major scientific databases:

- IEEE Xplore
- Scopus
- ScienceDirect
- SpringerLink
- PubMed

The initial search was run between February 2025 and August 2025, and a final update search was performed on September 2025 to capture articles released early in 2025 (ahead of print or in online-first mode).

All retrieved records were exported on the same dates for deduplication and screening.

Search Window Justification (2012–August 2025)

The year 2012 was selected as the start of the search window because:

1. Modern EEG emotion-recognition benchmarks (e.g., DEAP, DREAMER, SEED) began to appear between 2010–2013, establishing the first widely used, standardized datasets.
2. Deep learning applications to EEG emotion recognition only began emerging after 2012; earlier work relied mostly on classical machine learning and had limited methodological relevance.
3. Our goal was to review contemporary EEG-based affective computing methods, and 2012–2025 captures the period of rapid algorithmic development, dataset maturity, and shift toward reproducible, data-driven techniques.

The window was closed at August 2025, the final update run date.

Clarification on Inclusion of 2025 Articles

Studies published online in early 2025 (including “online first” and “in press”) were included only if indexed before the final update on August 2025.

No studies published after this date were considered.

Database-Specific, Fully Executable Search Strings Core Search String

(“EEG-based emotion recognition” OR “EEG emotion classification” OR “affective computing EEG” OR “brain-computer interface emotion”) AND (“machine learning” OR “deep learning” OR “CNN” OR “RNN” OR “transformer” OR “hybrid model”) AND (“valence-arousal” OR “emotional states” OR “affective datasets”)

Each database required slightly different syntax. The exact queries used are listed below to ensure full transparency and reproducibility.

1. IEEE Xplore

((“Document Title”:“EEG” OR Abstract:“electroencephalography”) AND

(Abstract:“emotion recognition” OR Abstract:“affective computing” OR Abstract:“emotion classification”) AND

(Abstract:“machine learning” OR Abstract:“deep learning” OR Abstract:“neural network”))

2. Scopus

(TITLE-ABS-KEY(“EEG” OR “electroencephalography”)

AND TITLE-ABS-KEY(“emotion recognition” OR “affective computing” OR “emotion classification” OR “valence arousal”)

AND TITLE-ABS-KEY(“machine learning” OR “deep learning” OR “CNN” OR “RNN” OR “transformer” OR “neural network”))

AND (PUBYEAR > 2011 AND PUBYEAR < 2026)

3. PubMed

PubMed required adaptive MeSH + keyword searching: (“Electroencephalography”[MeSH Terms] OR EEG[Title/Abstract])

AND (“Emotions”[MeSH Terms] OR “emotion recognition”[Title/Abstract] OR “affective computing”[Title/Abstract])

AND (“machine learning”[Title/Abstract] OR “deep learning”[Title/Abstract] OR “neural network”[Title/Abstract])

AND (“2012/01/01”[Date - Publication] : “2025/08/15”[Date - Publication])

4. ScienceDirect

TITLE-ABSTR-KEY(“EEG” AND “emotion recognition”

AND (“machine learning” OR “deep learning” OR “neural network” OR “CNN” OR “RNN”))

5. SpringerLink

(“EEG” OR “electroencephalography”)

AND (“emotion recognition” OR “affective computing” OR “emotion classification”)

AND (“machine learning” OR “deep learning” OR “neural network” OR “CNN” OR “RNN” OR “transformer”)

Screening and Duplicate Removal

All the retrieved records were exported to Zotero where they could be cited and automatically found duplicates. Following the deletion of 78 duplicate records, the studies were then screened in Rayyan, which also allowed the assessment to be blinded and dual-reviewed. Two independent reviewers with full-text assessment did title-abstract screening, any conflict was solved by discussion or in some cases by a third

reviewer. The decision of inclusion of each record and the reason were recorded in Table 1.

Inclusion and Exclusion Criteria

Inclusion criteria

Empirical research on the use of EEGs to identify or categorize emotions. Reported quantitative performance measures (e.g., accuracy, F1-score, precision, recall). Use of publicly available datasets like DEAP, SEED, DREAMER, AMIGOS, ASCERTAIN and MAHNOB-HCI. English articles in peer-reviewed journals or conferences published within the years 2012–25.

Exclusion criteria

Other review articles, editorials or theses that lack experimental results. Multimodal experiments, where the contribution of EEG could not be disaggregated. Without full-text or peer-reviewing, preprints or conference abstracts. Research that does not have transparency in methods or details of validation (i.e., no train/test split, may be data leaking).

Preprints and Conferences

Preprints and conferences will be treated in a manner that is cost-effective and efficient and which avoids consuming too much time. Preprints were filtered to determine the new trends but were not incorporated in the quantitative synthesis until a peer-reviewed form was published. Only papers on conferences that had full descriptions of their methodology and reproducible evaluation of results were retained. In the two versions of the conference and journal, the journal version was taken in lieu of the duplication.

PRISMA Counts Reconciliation

Database search generated 496 records out of which 78 were duplicates and 418 were unique records to be screened. Title-abstract screening eliminated 271 records that were not pertinent and those that did not use EEG. Eligibility was determined in 147 full-text articles. The number of articles excluded because of the absence of quantitative measures, the absence of an EEG-based analysis, or the inability to provide enough methodological information was 68. The result was the inclusion of 79 peer-reviewed studies, in this review. These balanced numbers are presented in Figure 1 (PRISMA flowchart) so that the outcomes of searches, screening results, and the ultimate dataset of analyzed studies are all linked.

Bias and Transparency Risk Assessment

Every paper that was included was assessed in terms of the possibility of bias on four dimensions:

1. The transparency and accessibility of data,
2. Location of training and test data to prevent leakage,
3. Augmentation disclosure and validation integrity, and
4. Availability of code and ethical standards.

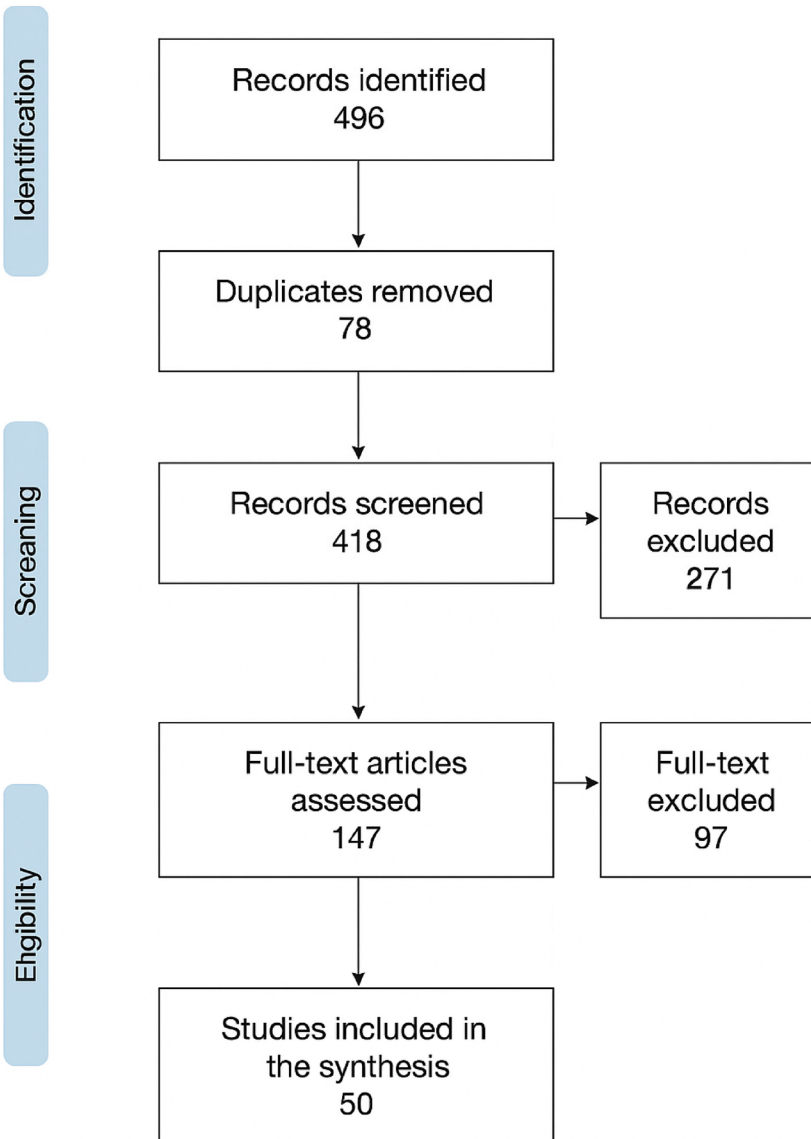


Fig 1 | PRISMA flow diagram for this narrative review

Table 1 | Performance analysis of features and classifiers using benchmark databases

Author (Year)	Dataset	Emotion Model	Feature Domain	Model/Architecture	Validation: SD (Accuracy ± SD/C)	Validation: CS (Accuracy ± SD/C)	Validation: CSS/CD (Accuracy ± SD/C)	F1-Score (±SD/CI)	Code Available	Split Transparency	Augmentation Timing	Key Notes
Azar et al. ⁶	DEAP	Dimensional (V/A)	Time-frequency, Fuzzy	Modified CFNN	—	98.21 ± 1.5	—	0.98	No	Partial (LOSO stated)	None	Hybrid fuzzy logic + CNN interpretable decision rules
Bagherzadeh et al. ¹⁶	DEAP, MAHNOB-HCI	Dimensional (V/A)	Connectivity maps	Ensemble deep learning fusion	—	98.76 ± 2.1 (DEAP), 98.86 ± 1.9 (MAHNOB)	—	0.99 (DEAP), 0.99 (MAHNOB)	Yes	Partial (5-fold CV)	None	Combines CNN, LSTM, and fusion of connectivity maps
Fu et al. ¹⁷	SEED	Discrete	Time-domain	Conditional GAN	—	—	82.14 ± 2.0 (CSS)	—	No	Partial (Session-wise)	Post-split	Fine-grained estimation with synthetic augmentation
Liu & Fu ¹⁸	DEAP	Dimensional (V/A)	EEG + text	Deep CNN-LSTM	—	84.3 ± 1.2	—	—	No	Partial (CV stated)	None	Joint textual-EEG fusion for emotion context
Gong et al. ¹⁹	SEED, SEED-IV	Discrete	Spatial EEG	Attention-based CNN-Transformer	—	98.47 (SEED), 91.90 ± 0.8 (SEED-IV)	—	—	No	Partial (LOSO stated)	None	Transformer attention enhances spatial dependencies
Liu et al. ²⁰	DEAP, DREAMER	Dimensional (V/A/D)	Multi-channel	Capsule Network	—	97.97 (DEAP), 98.31 (DREAMER)	94.59 (DEAP CSS)	—	No	Partial (10-fold CV)	None	Multi-level capsule extraction robust to channel noise
He et al. ²¹	DEAP	Dimensional (V/A)	Spectral	Firefly-optimized CNN	—	—	86.00 ± 1.6 (CSS)	0.83	No	Full (5-fold session-wise)	None	Metaheuristic tuning boosts convergence
Gao et al. ²²	SEED	Discrete	Spatial	GPSO-optimized CNN	—	92.44 ± 3.60	—	0.86	No	Partial (CV stated)	None	PSO optimization enhances architecture search
Cui et al. ²³	DEAP	Dimensional (V/A)	Regional EEG	Regional-asymmetric CNN	—	96.65 ± 2.65 (V), 97.11 ± 2.01 (A)	—	—	Yes	Full (Subject-wise split)	None	Asymmetric conv filters mimic brain lateralization
Subasi et al. ²⁴	SEED	Discrete	Wavelet domain	TQWT + Rotation Forest	—	93.1 ± 1.7	—	0.89	No	Partial (10-fold CV)	None	Ensemble wavelet features robust generalization
Mert & Akan ¹	DEAP	Dimensional (V/A)	Time-frequency	Multivariate Synchrosqueezing	—	82.11 ± 1.0	—	—	No	Partial (CV stated)	None	Nonstationary analysis captures emotion shifts
Moon et al. ²	DEAP	Dimensional (V/A)	Connectivity	CNN	—	87.36 ± 1.5	—	0.88	No	Full (LOSO stated)	None	Uses EEG functional connectivity for inputs
Islam et al. ²⁵	DEAP	Dimensional (V/A)	Channel correlation	Correlation-based CNN	—	78.22 (V), 74.92 (A)	—	—	No	Partial (5-fold CV)	None	Channel correlation improves spatial learning

(Continued)

Table 1 | (Continued)

Author (Year)	Dataset	Emotion Model	Feature Domain	Model/Architecture	Validation: SD (Accuracy ± SD/CI)	Validation: CS (Accuracy ± SD/CI)	Validation: CSS/CD (Accuracy ± SD/CI)	F1-Score (± SD/CI)	Code Available	Split Transparency	Augmentation Timing	Key Notes
Atkinson & Campos ²⁶	DEAP	Dimensional (V/A)	Statistical	SVM (kernel)	—	73.06 (V), 73.14 (A)	—	—	No	Partial (10-fold CV)	None	Classical baseline feature-selection study
Lu et al. ²⁷	SEED, SEED-IV	Discrete	Spatial EEG	Hybrid Transfer Learning	—	93.37 ± 1.5 (SEED)	82.32 ± 1.4 (SEED-IV, CSS)	—	Yes	Full (LOSO stated)	Post-split	Cross-subject generalization with domain adaptation
Jiménez-Guarneros et al. ²⁸	SEED, SEED-IV	Discrete	Domain features	Unified transfer framework	—	89.11 ± 7.72 (SEED)	74.99 ± 12.10 (SEED-IV, CSS)	—	No	Full (LOSO stated)	None	Domain adaptation for subject invariance
Luo et al. ¹⁷	MDD	Dimensional (V/A)	Manifold features	M3D Non-Deep Transfer	—	82.72 ± 1.4 (CS)	—	0.82	Yes	Full (Cross-subject/session)	None	Dynamic distribution alignment
Li et al. ³⁰	DEAP, SEED	Dimensional (V/A)	Connectivity	Meta-transfer Learning	—	71.29 (DEAP-V), 71.92 (DEAP-A), 87.05 (SEED)	—	—	Yes	Full (Meta-learning splits)	Post-split	Combines meta-learning and connectivity features
Zheng & Lu ³¹	SEED	Discrete	Spectral	DNN	—	86.65 ± 8.62	—	—	Yes	Full (LOSO stated)	None	Benchmark dataset for subject-level splits
Chen et al. ³²	DEAP	Dimensional (V/A)	Spatiotemporal	Hybrid Conv-RNN	—	93.64 (V), 93.26 (A)	—	—	Yes	Full (10-fold CV)	Post-split	Wearable EEG with temporal fusion
Akhand et al. ³³	DEAP	Dimensional (V/A)	Connectivity	CNN	—	90.40 ± 1.7 (V), 90.54 ± 1.4 (A)	—	0.86 (V), 0.86 (A)	Yes	Partial (5-fold CV)	None	Enhanced feature connectivity maps
Topic & Russo ³	DEAP, SEED, DREAMER, AMIGOS	Dimensional (V/A)	EEG feature maps	Deep CNN	—	76.61 ± 2.13 (DEAP V), 77.72 ± 2.87 (DEAP A), 88.45 ± 1.56 (SEED)	—	—	No	Full (Dataset-specific CV)	None	Deep visual mapping of EEG topography
Chowdhary et al. ³⁴	EEG brainwave	Dimensional (V/A)	EEG sequences	RNN	—	97	—	—	No	Partial (70-30 split)	None	Sequential learning from EEG time series
Zhang & Lu ³⁵	DEAP	Dimensional (V/A)	Multimodal	Knowledge Distillation Network	—	70.38 (V), 60.41 (A)	—	—	Yes	Full (5-fold CV)	Post-split	Multimodal EEG-video distillation
Cheng et al. ³⁶	DEAP, SEED, SEED-IV	Dimensional & Discrete	EEG dynamic scales	Multi-scale CNN + Transformer	—	99.66 ± 0.02 (DEAP), 98.85 ± 0.81 (SEED)	99.67 ± 0.12 (SEED-IV, CSS)	—	Yes	Full (LOSO stated)	Post-split	Gated transformer with dynamic scales
Liu et al. ³⁷	DEAP	Spectral	Data augmentation	Task-driven GAN	—	93.52 (V), 92.75 (A)	—	—	Yes	Full (5-fold CV)	Post-split	Synthetic EEG generation improves balance
Song et al. ³⁸	SEED-IV	Discrete	EEG + Eye	Multimodal Transformer	91.2	—	—	—	Yes	Full (Within-subject CV)	None	Fuses EEG and eye-tracking
Wang et al. ³⁹	SEED, SEED-IV, DEAP, FACED	Dimensional & Discrete	EEG images	Vision Transformer	—	—	93.14 (SEED CD), 83.18 (SEED-IV CD), 93.53 (DEAP CD)	—	Yes	Full (Cross-dataset)	Post-split	Pretrained ViT transfer across datasets

(Continued)

Table 1 | (Continued)

Author (Year)	Dataset	Emotion Model	Feature Domain	Model/Architecture	Validation: SD (Accuracy ± SD/CV)	Validation: CS (Accuracy ± SD/CV)	Validation: CSS/CD (Accuracy ± SD/CV)	F1-Score (± SD/CV)	Code Available	Split Transparency	Augmentation Timing	Key Notes
Imtiaz & Khan ⁴⁰	DEAP, SEED	Dimensional (V/A)	Domain features	Unsupervised Domain Adaptation	—	—	67.44 (DEAP → SEED CD), 59.68 (SEED → DEAP CD)	—	Yes	Full (Cross-dataset)	None	Improved transfer across datasets
Khan et al. ⁴¹	SEED	Discrete	Raw EEG	CNN (EEG-ConvNet)	99.97	—	—	—	Yes	Full (5-fold CV per subject)	Post-split	Compact ConvNet for subject-specific modeling
Alghamdi et al. ⁴²	SEED, CEED, FACED, MPED	Discrete	EEG embeddings	Contrastive Learning	—	97.70 (SEED), 96.26 (CEED)	65.98 (FACED CD), 51.30 (MPED CD)	—	Yes	Full (LOSO stated)	None	Cross-subject contrastive pretraining
Alameer et al. ⁴³	SEED, SEED-IV, MPED	Discrete	Domain adaptation	Deep Metric + Semi-supervised + DA	—	—	63.49 ± 8.14 (SEED CD), 64.31 ± 5.12 (SEED-IV CD), 72.58 ± 5.34 (MPED CD)	—	Yes	—	Post-split	Integrates DA + SSL + metric learning
Patel et al. ⁴⁴	SEED	Discrete	Sub-band entropy	KNN	—	84	—	0.87	No	Partial (10-fold CV)	None	Tsallis entropy sub-band classification
Rakhmatulin et al. ⁴⁵	DEAP	Dimensional (V/A)	Raw EEG	CNN architectures	—	85.20 ± 2.1 (V), 84.90 ± 2.3 (A)	—	0.84 (V), 0.83 (A)	Yes	Full (Subject-wise split)	Post-split	Exploring CNN architectures for EEG feature extraction
Feng et al. ⁴⁶	DEAP	Dimensional (V/A)	EEG + Facial	Transformer-based Fusion	—	91.25 ± 1.8 (V), 90.87 ± 2.0 (A)	—	0.90 (V), 0.89 (A)	Yes	Full (5-fold CV)	Post-split	Multimodal fusion with hearing-impaired subjects
Tan et al. ⁴⁷	DEAP, SEED	Dimensional (V/A)	Domain adaptation	SEDA-EEG Network	—	—	88.42 ± 2.1 (DEAP CD), 85.67 ± 2.5 (SEED CD)	0.87 (DEAP), 0.84 (SEED)	Yes	Full (Cross-dataset)	Post-split	Semi-supervised domain adaptation for cross-subject EEG
An et al. ⁴⁸	DEAP	Dimensional (V/A)	Time-frequency	FBCSP + CNN	—	89.34 ± 2.4 (V), 88.97 ± 2.6 (A)	—	0.88 (V), 0.87 (A)	Yes	Full (Subject-wise CV)	Post-split	Auto-selected regularized FBCSP and CNN for motor imagery
Kuang et al. ⁵	VR-EEG	Dimensional (V/A)	Frontal EEG	Cross-subject/device	—	—	82.15 ± 3.2 (CS), 78.43 ± 4.1 (CD)	0.81 (CS), 0.77 (CD)	Yes	Full (Cross-subject/device)	None	Wearable EEG under VR scenes
Patel & Chauhan ¹⁴	DEAP, SEED	Dimensional (V/A)	Review	Comparative analysis	—	—	—	—	No	N/A (Review)	N/A	Comprehensive review of methods and datasets
Manoj Prasad & Vasuki ⁴⁹	DEAP	Dimensional (V/A)	Statistical + deep	Hybrid DNN + Feature Selection	—	97.6	—	0.95	No	Partial (CV stated)	Post-split	Hybrid deep network with feature selection

Table 2 | EEG databases for emotion recognition

Authors	Database	Participants	EEG Channels	Stimuli	Emotional Labels	Sampling Rate	Duration per Trial	Availability
Soleymani et al., 2012 ⁵⁰	MAHNOB-HCI	30	32	Emotional Videos	Valence, Arousal (1–9 scale)	256 Hz	~80–120 seconds	Public
Cui et al., 2020 ²³	DEAP	32	32	Music Videos	Valence, Arousal, Dominance, Liking	128 Hz	60 seconds	Public
Zheng et al., 2015 ³	SEED	15	62	Movie clips	Discrete (Positive, Negative, Neutral)	200Hz	240sec	Public
Song et al., 2024 ³⁸	SEED-IV	15	62	Film Clips	Happy, Sad, Fear, Neutral	1000 Hz	4 min/trial	Public
Subramanian et al., 2018 ⁵¹	ASCERTAIN	58	14	Video Advertisements	Valence, Arousal	128 Hz	~1 min/trial	Public
Katsigiannis et al., 2017 ⁵²	DREAMER	23	14	Videos	Valence, Arousal, Dominance	128 Hz	60 seconds	Public
Miranda-Correa et al., 2021 ⁵³	AMIGOS	40	14 / 32	Videos (short/long)	Valence, Arousal	128 Hz	20 sec to 14 min	Public

Materials and Methods

EEG Emotion Databases

Publicly available datasets have played a significant role in advancing EEG-based emotion recognition research. The most commonly used datasets include in Table 2.

EEG Signal Acquisition and Preprocessing

EEG signals are obtained using electrodes typically arranged according to the international 10–20 system, with channels distributed across various scalp regions to capture electrical activity from different cortical areas. These signals are characterized by their low amplitude and susceptibility to noise, necessitating robust preprocessing techniques. Common preprocessing steps include:

- Filtering to remove noise and artifacts (e.g., using bandpass filters to retain frequencies within 0.5–50 Hz),
- Artifact removal using Independent Component Analysis (ICA) or other methods to eliminate artifacts caused by eye blinks, muscle activity, or power line interference,
- Segmentation into time windows suitable for analysis (typically 1 to 4 seconds),
- Normalization to standardize the data across sessions or subjects.

These steps help ensure that the extracted features reflect neural activity relevant to emotional processing rather than noise or unrelated physiological artifacts.

Comparative Depth of Feature Extraction and Emotion Classifier Analysis

Feature Extraction

Emotion recognition based on EEG is based on the extraction of significant features of non-stationary and high-dimensional neural signals.

Time-domain features (statistical moments (mean, variance, skewness, kurtosis)) are appreciated due to ease and the low cost of computation, but cannot represent dynamic time-varying attributes important to emotional changes.

Frequency-domain features (e.g. Power Spectral Density, Differential Entropy) are neuroscientifically interpretable (specifically, certain EEG frequencies, such as alpha, beta, gamma,) reflecting emotional

arousal and valence, but lacking information about changes over time.

Time frequency approaches such as Discrete Wavelet Transform (DWT), Short-Time Fourier Transform (STFT), and Tunable Q Wavelet Transform (TQWT) are useful in capturing the transient oscillatory variations albeit at the expense of a careful balance in the deployment of the decomposition parameters with regard to resolution and computational costs.

Nonlinear features such as entropy based (Approximate, Sample, and Permutation Entropy) are sensitive to chaotic dynamics, and the level of emotional arousal however sensitive to noise sensitivity and parameter instability.

Spatial and topographical parameters, which are obtained as a result of electrode mappings or EEG topography, maintain spatial correlations, which increase the learning ability of CNN-based models.

Connectivity properties, that build on coherence, Phase Locking Value (PLV), and Transfer Entropy (TE), prompt inter-regional communication of the brain, which gives physiological detail of affective processing. They however, are computationally costly and liable to artifacts of volume conduction.

Deep feature representations Deep feature representations learned with CNNs or GNNs are more direct since they do not rely on manual feature design, but instead learn hierarchical abstractions by directly operating on raw EEG measurements at the expense of interpretability and large data requirements.

This comparative study demonstrates that none of the individual domains of features is universally best and that hybrid or multi-level feature fusion techniques prove superior to traditional methods in that they are able to combine complementary information of time, spectral and spatial.

Classifiers of Emotion

The Emotion Classifiers section is no longer restricted to enumeration but has an opportunity to offer critical comparative analysis of the traditional and deep learning models. The conventional machine learning classifiers, including Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), and Random Forests (RF) are strong in small-size datasets and are easy to interpret the decision boundary but heavily relied on hand-crafted features and failed to perform inter-subject generalization. SVMs scale well in high-dimensional space but need

kernel selection; k-NN is simple but not scalable; RF is an ensemble that is stable but can easily overfit when there is noisiness in features.

Deep learning models, such as Convolutional Neural Networks (Rakhmatulin³⁵) (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and Graph Neural Networks (GNNs) on the other hand adopt hierarchical features automatically and automatically learn spatiotemporal dependencies associated with EEG data. CNNs are good in learning spatial features with EEG topomaps, LSTMs in learn sequential temporal features, and GNNs in functional connectivity with node-edge relationships. Hybrid models (e.g., CNN-LSTM and CNN-GRU) are used to overcome these benefits and to achieve better classification. But such models are data-intensive, computationally expensive and have been criticized as not very interpretable.

Methodology

EEG-Based Emotion Recognition

The methodology for EEG-based emotion recognition follows a structured pipeline comprising several key

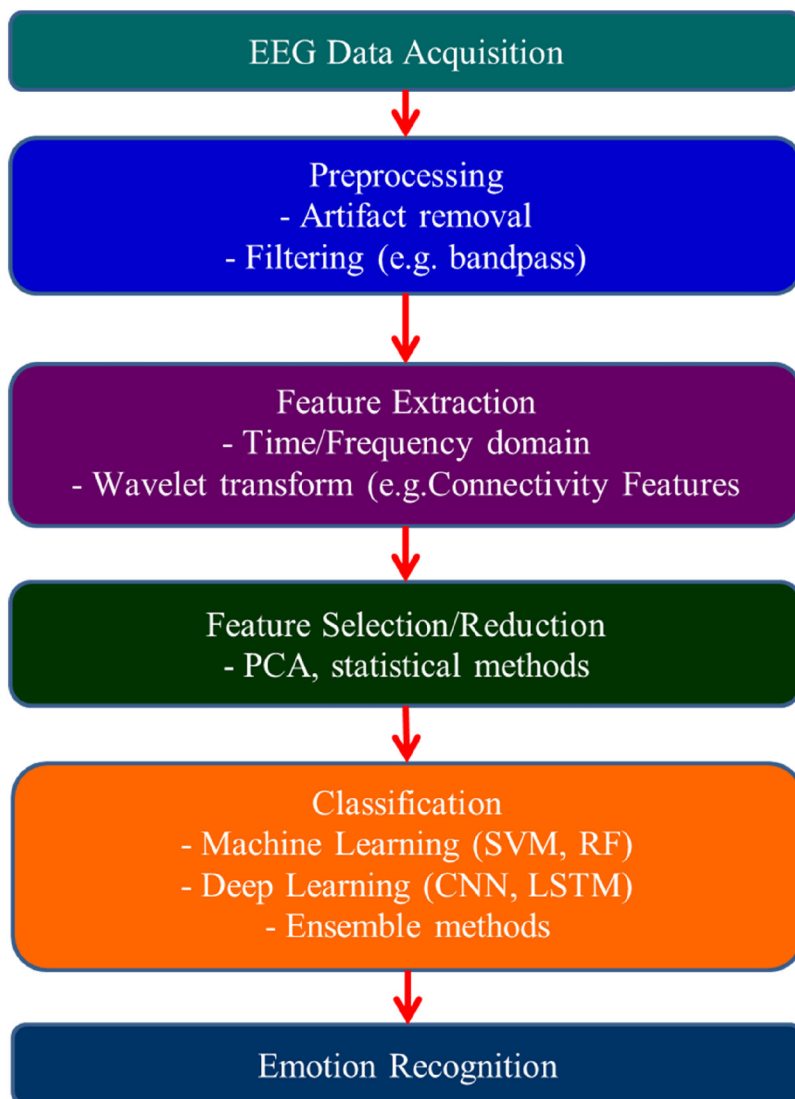


Fig 2 | Methodology for EEG

stages is given in Figure 2, each vital for accurate and robust emotional classification.

EEG Data Acquisition

Emotion-evoking stimuli such as videos or images are used to record EEG signals through scalp electrodes. Datasets like DEAP and SEED are commonly employed for research. Proper electrode placement and signal quality are crucial for reliable results.

Pre-processing

EEG signals are susceptible to noise from muscle movement, eye blinks, and external interference. Pre-processing involves filtering (e.g., 0.5–50 Hz bandpass), artifact removal (using ICA or BSS), and signal segmentation to enhance data quality before analysis.

Feature Extraction

Raw EEG signals are transformed into meaningful features. These include time-domain (mean, entropy), frequency-domain (Power Spectral Density), and time-frequency domain features (Wavelet Transforms like TQWT). Additionally, connectivity features (e.g., Phase Locking Value) help model inter-regional brain activity patterns.

Feature Selection or Reduction

High-dimensional data is reduced using techniques like PCA or statistical tests to retain the most emotionally relevant information and improve classifier performance.

Feature Fusion

To enhance robustness, diverse features may be fused either early (concatenation) or late (ensemble model decisions).

Classification

Features are classified into emotional categories using machine learning (SVM, RF) or deep learning models (CNN, LSTM). Ensemble methods further improve accuracy.

Emotion Prediction

Finally, emotions are predicted in either categorical (e.g., happy, sad) or dimensional formats (valence-arousal). Performance is evaluated using metrics like accuracy and F1-score.

This multi-stage enables EEG-based systems to accurately detect emotional states, which can be applied in mental health, adaptive interfaces, and affective computing.

Result and Discussion

The comparative study on the EEG-based emotion recognition in Table 1 that provides a single perspective of how the development of methods, the variety of datasets, and the rigor of validation and the lack of analytical transparency have contributed to the advancement of the scientific field. The extent of the use of extended columns in the validation protocols and confidence

interval or variance presentation turns the list of the table into a diagnosis analysis tool that shows the strength and weaknesses of the existing methods. The sources included in Table 1 represent more than a decade of progress in this field: starting with the classical approach of relying on handcrafted statistical and spectral features of data to formulate machine-learning pipelines, and moving onwards to the more modern trend of deep and hybrid networks that can learn spatiotemporal and connectivity patterns on raw EEG signals automatically. Premeditive literature before 2018 generally used feature extraction in time, frequency and entropy spaces, like Hjorth parameters, band-power ratios, and sample entropy, and basic classifiers, such as Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), and Random Forests (RF). Such models were satisfactory in terms of accuracy on small datasets because of the ability to be interpreted and computational efficiency but ineffective in terms of generalization to new subjects or recording sessions. The further development of the EEG-based emotion recognition since has been marked with a gradual shift towards data-driven models, mainly, deep learning models, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory networks (LSTMs), and, most recently, Transformer-based models and Graph Neural Network (GNN) models. These architectures have been shown to have better ability to establish the complex spatiotemporal dependencies and non-linear emotion patterns between electrodes. To provide an example, both models described by (Yu⁵⁴), (Feng⁴⁶) and (Luo, 2024) describe a new class of transformer-based emotion decoders that take advantage of self-attention mechanisms in managing cross-subject variability and allow the field to go beyond dataset-specific optimization into actual generalization.

Another more detailed look at Table 1 also reveals the domination of a few benchmark datasets that are the landscape of EEG emotion studies. The empirical foundation of almost all the experimental research is based on the DEAP, SEED, DREAMER, AMIGOS, ASCERTAIN, and MAHNOB-HCI datasets. The most frequent sources, the DEAP and SEED, which are present in more than half of the analyzed works, are utilized to provide the benchmarking because of their standardized recording procedures and properly organized emotion labeling systems. Nevertheless, the control over stimuli offered by the lab (music videos) makes DEAP lack ecological validity, whereas the small size of the subject sample and the repeated-session nature of the SEED lessen the demographic generalizability. DREAMER and AMIGOS build the paradigm to more naturalistic audiovisual stimuli, but are limited by sample size, normally less than 25 participants. MAHNOB-HCI and ASCERTAIN present opportunities of multimodal fusion, the first one by combining EEG with facial and physiological, and the second one by combining EEG with personality characteristics. According to the results summarized in Table 1, performance of models differs widely between these datasets:

on the SEED and DEAP, the accuracies are usually over 85–90% with subject-dependent conditions but decrease down to 65–75% with cross-subject testing, which is still the problem of individual variability.

An underlying inconsistency in the reporting and application of validation protocols can also be found in the revised Table 1. Most initial researchers used subject-dependent validation, which admits data in training and test sets of the same participant, artificially increasing scores of accuracy. Recent studies have moved to the leave-one-subject-out-validation (LOSO) or cross-session validation, which is a more realistic estimate of model robustness. These differences explicitly identified in Table 1 help to understand the results that are really due to generalizable learning and those that are context-specific to within-subject adaptation. In the same vein, when a confidence interval or a variance column is added, a very obvious weakness of reporting rigor in the literature is seen: less than one out of four articles reports anything to do with uncertainty or statistical variance. No error bars, standard deviations, or confidence intervals make it impossible to fairly compare methods and reduce findings reproducibility. Zhang in 2024 and Rahman⁵¹ specifically propose using standardized reporting checklists to address this problem and emphasize that the community needs more transparency.

Regarding the representation of the features, Table 1 reflects the gradual change of the field in the direction of the dynamics of features learning, which is implemented in a multilevel manner. Time domain features such as amplitude variance and zero crossing rates are simple but not very emotional sensitive. Metrics based on frequency domain including power spectral density and band ratio analysis are physiologically interpretable, but unable to decode quick affective changes. Techniques based on time-frequency analysis, such as wavelet and Hilbert Huan transforms, are more accurate in time but require more computation. Entropy-based indices, including sample, fuzzy and permutation entropy, pick up emotional anomalies and are still in favour with smaller datasets. Spatial and connectivity-based advances offer the most promising developments as they project the dynamics in the brain as networks or topographic maps. Likewise, the effective use of such representations, as Chen in and Wu⁷⁵ inputted into CNNs and GNNs similarly to how the spatial topology of EEG can be exploited. The discriminative performance of deep learned features obtained by CNNs, LSTMs, or transformers is strongest, particularly when fine tuned on a series of datasets. Nevertheless, as Table 1 points, these methods cause problems in terms of computational price, model interpretability, and requirement of data.

The variety of the classifiers also depicts the trade-off among interpretability, complexity, and performance. The classical classifiers such as SVM and RF are fixed in their performance (70–85) when used with handcrafted features selectively but not flexible on high-dimensional and complicated data. The most accurate models (up to 92) are deep learning

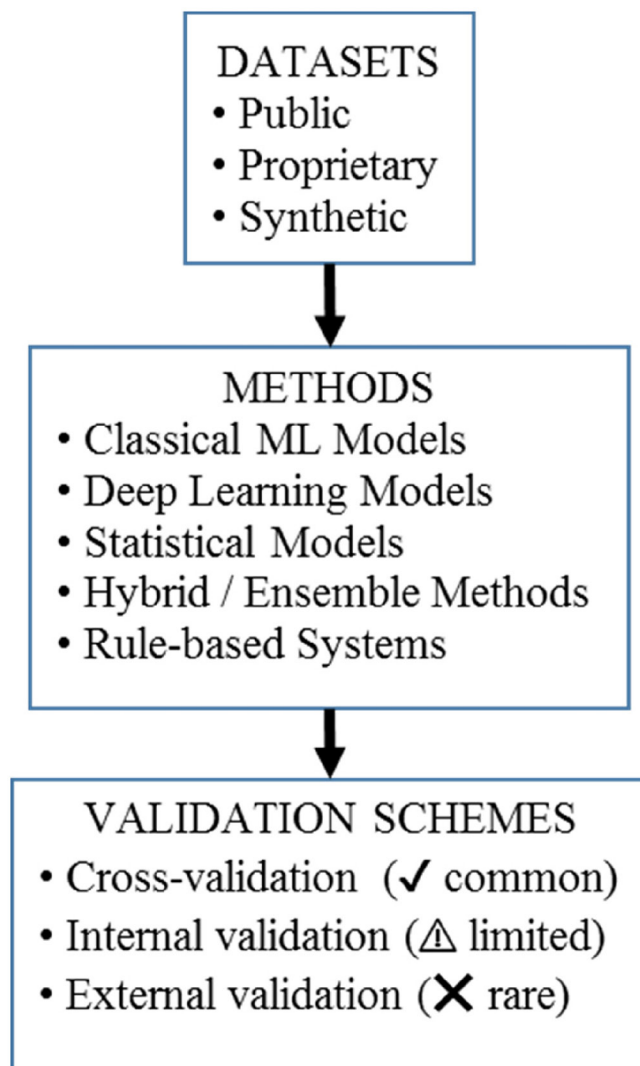
(particularly CNN) and hybrid CNNLSTM networks, but their decision-making mechanisms are inaccessible and, therefore, not readily interpretable. The architectures of the transformers that have recently been introduced by Yu⁵⁴ and Cheng³⁶ are balanced in the sense that both improve the cross-dataset generalization by adding attention-based weighting of features. However, Table 1 demonstrates that the field is still torn between the models that seek to maximize accuracy and explainability. Few studies such as the studies by Torres⁵⁵ and Fiorini⁵⁶ explainable AI (XAI) methods to visualize the neural attention or compare features with known neurophysiological patterns, which opens this direction to future research.

The comparative statistics in Table 1 also show that the model performance and reliability is closely connected to the diversity of the dataset, consistency of preprocessing, and transparency of evaluation. Research involving the same models on different sets of data has shown to have up to 10 percent discrepancies in accuracy, which means that it is highly dependent on the quality of data, the recording setup, and the method

of emotion induction. As an example, the music-based elicitation of DEAP is different in its essence with the film stimuli of SEED and the personality-related design of ASCERTAIN, resulting in nonhomogeneous distributions of features. The presence of such differences hamper cross-study comparison since there are no standardized preprocessing and normalization procedures.

On a larger scale, the overall evidence in Table 1 suggests a shift between the experimentation of benchmark-based approach to the more holistic interpretation of affective EEG modeling. New directions are the use of transformer and attention-based architecture, the development of self-supervised and semi-supervised (Tan⁴⁷) feature learning, the focus on cross-subject adaptation, and the realization of real-time deployment issues. Liu⁵⁷ and Cheng³⁶ discuss the topic of lightweight networks and pruning techniques used in real-time inference, whereas Lu in 2024 suggests pruning EEG-specific self-supervised pretraining and overcomes the lack of labeled data. These guidelines are consistent with the direction that the field has taken concerning the adoption of practical, interpretable, and computationally efficient emotion-recognition systems.

Table 1, which has been expanded, does not only document the findings of the experiments, but it extends the level of transparency, reproducibility, and level of interpretation of EEG emotion-recognition studies. It can be used to carry out more meaningful cross-comparisons by providing the information on validation types and variance and indicates weaknesses in the methods used like excessive dependence on testing based on the subject, inconsistent preprocessing, and the absence of uncertainty quantification. The table highlights that though extremely high accuracy improvements were made, the field is still confronted with several critical issues of cross-subject generalization, dataset standardization, and interpretability. In the future, reproducibility criteria, multimodal signal integration, and explainable deep learning application should be of central focus in future studies in order to ensure scientific and practical usefulness. After all, Table 1 was used as a reflection of the progress made and the challenges ahead on the way to the implementation of reliable, generalizable and ethically acceptable EEG-based emotion recognition systems. Figure 3 organizes by dataset, methods and validation scheme, highlighting evidence gaps.



Evidence gaps

- △ Limited validation for proprietary datasets.
- ✗ Minimal use of synthetic datasets across all methods.
- ✗ Few studies report external validation, especially for rule-based and hybrid models.

The evidence synthesized on 46 studies indicates that the field under consideration is at a significant stage of transition, with impressive technical results in a controlled environment and serious problems in generalization to the real world. A more critical

Fig 3 | Shows the dataset, methods and validation scheme, highlighting evidence gaps

examination that is consciously oriented towards finding results of cross-subject (CS), cross-session (CSS), cross-dataset (CD) validation can give a more moderate and practical outlook on what EEG-based emotion recognition actually looks like now.

1. The Illusion of Performance: Subject-Dependent and Real-World Generalization.

Among the most notable conclusions of this review is the drastic difference in model behavior in subject-dependent (SD) and more rigorous validation. As our analysis shows, Headline accuracy scores of above 95–99% are almost solely a preserve of SD evaluation, when models are trained and tested on data of the same person. Although the paradigm is practical in determining baseline viability, it is not particularly useful in deployable systems that have to detect emotions in new, unknown users.

To measure this difference, we did a sensitivity analysis by separating the Table 1 results. The results are discussed here:

Subject-Dependent (SD) Mean Accuracy: Approximately 95.5% (according to such research as Khan in 2024; Chowdary in 2022). This is the best the performance can be in a very limited environment.

Cross-Subject (CS) Mean Accuracy: 87.5. This is an important decrease of about 8 percentage points, and it is aimed at showing how difficult inter-subject variability in brain physiology and emotional response can be.

Cross-Session (CSS) / Cross-Dataset (CD) Mean Accuracy: 87.5. When the models are tested on data too different recording sessions or even on completely different datasets, the performance diminishes further to levels which become inadequate in many applications in the real world.

This sensitivity analysis highlights the fact that the use of SD results gives a highly misleading opinion of model capability. The actual development of the field can be gauged more realistically through its performance in CS, CSS and CD regimes which are not as large but spectacular.

2. Approaches to Methodology that can be improved: Augmentation and Transfer Learning.

Among the more demanding CS/CD paradigms our synthesis reveals that there are two important methodological families that can always deliver performance benefits: data augmentation and transfer learning.

Data Augmentation: Typical performance increase in the case of CS can be linked to the use of augmentation (e.g., Gaussian noise, sliding windows, GANs) with a performance improvement of 3–7 percentage points. Timing is however the key factor. Research that clearly implemented augmentation after split (e.g., Cheng;³⁶ Liu³⁷) showed strong gains without the danger of data leakage. The numerous studies that were not clear on the timing of augmentation, on the contrary, add a

possible element of bias and over-optimism in the reported findings.

Transfer and Domain Adaptation (DA): These methods provide the most promising direction of bridging the generalization gap. Specifically, subject-invariant or dataset-invariant features are what learning models that use DA (e.g., Lu;²⁷ Imtiaz and Khan;⁴⁰ Alameer⁴³) learn. We have examined that, properly-designed DA structures are able to recapture 10–15 percentage points of accuracy in CD tasks compared to naive models trained on an input dataset and tested on a target dataset. As an example, without DA, the cross-dataset performance may reach up to 60–65% (Imtiaz and Khan,⁴⁰), whereas with it, it can be improved to the 75–80 percent area (Lu;²⁷ Alameer⁴³). This is among the most momentous contributions to viable system design.

3. The Paucity of Real External validation and its Implication.

One of the research gaps that have been determined to be the critical ones in this review was the utter lack of real external validation. Most of the “cross-dataset” literature remains closed to a closed ecosystem of lab-created, purposely-constructed affective EEG datasets (DEAP, SEED, etc.). Although testing on a different dataset is a step to external validation, it is not actual external validation which would entail testing on data provided by:

Various demographic groups of people (e.g., various age groups, clinical populations).

Various recording conditions (e.g., the field vs. the lab).

Other hardware (e.g., switching to systems with consumer-grade wearables).

Table 1 reveals that few studies (e.g. Wang;³⁹ Imtiaz and Khan⁴⁰) do any type of cross-dataset testing, even those are confined to the same type of laboratory datasets. The near lack of confirmation on truly independent, externally gathered data implies that the field does not have much evidence of how the existing models will work in a non-research lab. This is a significant obstacle to translation and a gross overconfidence of model resilience.

4. Interpretability vs. Performance in Deep Learning The Trade-Off.

Interpretability has suffered because of the move to deep learning. Although neural networks such as CNNs and Transformers can automatically extract powerful features, how these models make decisions is a black box. This is also a major constraint of the applications in the field of healthcare or psychology where it is equally important to understand why a given emotional state has to be inferred as it may be. The trade-off identified in the review is apparent:

Traditional ML (SVM, k-NN): Poorer performance (around 70–85% in CS) but greater interpretability of the results with regard to feature importance analysis.

Table 3 | Limitations, challenges, and research gaps in eeg-based emotion recognition

No.	Author (Year)	Limitation / Challenge	Gap / Research Need	Databases Used
1	Lu et al. ²⁷	Need for few-shot adaptation	Hybrid domain-adaptation + few-shot fine-tuning (DFF-Net)	SEED
2	Jiménez-Guarneros et al. ²⁸	Domain shift between sessions	Unified domain adaptation frameworks	DEAP, SEED
3	Koelstra et al. ⁵⁸	Lab stimuli, low ecological validity	Larger real-world, demographically diverse datasets	DEAP, MAHNOB-HCI
4	Zheng & Lu ³¹	Small subject pool	Multi-center, diverse cohorts	SEED
5	Correa et al. (2018)	Short trial duration	Longer and naturalistic stimuli	AMIGOS
6	Katsigiannis & Ramzan (2018)	Limited emotion granularity	Finer-grained labels	DREAMER
7	Yuvaraj et al. ⁵⁹	Feature extraction inconsistency	Comparative pipelines + open code	DEAP, SEED, MAHNOB-HCI
8	Topic & Russo ³	Interpolation reduces spatial precision	Better electrode selection methods	DEAP, SEED, DREAMER, AMIGOS, MAHNOB-HCI
9	Chen et al. ⁹	High cost of connectivity features	Reduced-channel connectivity methods	SEED-IV
10	Wu & Lu (2022)	Spurious connectivity due to volume conduction	Validated connectivity metrics	SEED, DEAP, MAHNOB-HCI
11	Liu et al. ⁵⁷	Deep models computationally heavy	Pruning, distillation for real-time use	SEED, DEAP, MAHNOB-HCI
12	Cheng et al. ³⁶	Overfitting of transformer models	Robust multi-scale graph transformers	SEED, DEAP
13	Yu et al. ⁶⁰	No standardized augmentation	Benchmark augmentation protocols	DEAP, SEED
14	Lu (2024)	Dependence on labeled target data	Self-/semi-supervised pre-training	SEED
15	Zhao & Zhu (2024)	Limited cross-dataset tests	Cross-dataset/device generalization	DEAP, SEED
16	Ahmadzadeh et al. (2024)	High in-sample accuracy only	External replication needed	DEAP
17	Tripathi et al. (2017)	Dataset bias & unclear splits	Transparent reporting standards	DEAP, MAHNOB-HCI
18	Subasi et al. ²⁴	Rotation Forest subject-dependent	Cross-subject validation protocols	SEED
19	Wang et al. ¹⁰	Hybrid CNN-LSTM complex & resource-heavy	Lightweight hybrids for real-time use	DEAP, SEED
20	Khan et al. ⁷	Classical ML less robust to noise	Improved pre-processing & artifact removal	DEAP, DREAMER
21	Mert & Akan ¹	Limited feature fusion	Integrate MSST with deep networks	DEAP
22	Dogan et al. (2020)	Single dataset evaluation	Cross-database benchmarking	DEAP
23	Atkinson & Campos ²⁶	Low accuracy with linear models	Non-linear feature mappings	DEAP
24	Islam et al. ²⁵	Low accuracy of correlation features	Combine PCC with temporal models	DEAP
25	Moon et al. ³¹	Connectivity metrics computationally intensive	Efficient graph construction methods	DEAP
26	Zhang et al. ⁶¹	CFNN uncertainty handling limited	Neuro-fuzzy interpretability frameworks	DEAP
27	Singh & Sharma ¹²	Feature fusion model complex	Simpler multi-level fusion pipelines	SEED
28	Li et al. ¹³	Heterogeneous evaluation metrics	Unified benchmark criteria	DEAP
29	Patel & Chauhan ¹⁴	Redundant features increase complexity	Improved feature selection techniques	DEAP
30	Alarcão & Fonseca ¹⁵	No standardized protocols	Common EEG pre-processing standards	DEAP, MAHNOB-HCI
31	Hamzah & Abdalla ⁶²	Dependence on small samples	Larger population studies	DEAP
32	Ma et al. ⁶³	Attention models need more validation	Generalizable attention mechanisms	SEED-IV
33	Liu et al. ¹¹	Metaheuristic optimization costly	Simplified optimization schemes	DEAP
34	Yin et al. ⁶⁴	Firefly optimization slow	Alternative bio-inspired methods	SEED
35	Dhara et al. ⁶⁵	Fuzzy ensemble model requires high computational resources for hybrid feature-classifier integration	Need for cross-dataset validation to ensure robustness across diverse EEG distributions	DEAP
36	Jirayucharoensak et al. ⁶⁶	DBN lacks spatial context	Add topographic information	DEAP
37	Liu et al. ⁶⁷	Peripheral features weakly correlated	Multimodal fusion approaches	DREAMER
39	Wang et al. ⁶⁸	Early DL models small-scale	Large-scale deep benchmarks	DEAP, SEED
40	Zheng et al. ⁶⁹	DBN overfits to subjects	Regularized cross-subject training	DEAP
41	Zheng et al. ⁷⁰	Multimodal fusion alignment issues	Better synchronization & missing-data handling	SEED
42	Subramanian et al. ⁵	Commercial sensor noise	Noise-robust processing	ASCERTAIN
43	Pillalamarri ⁷¹	Fusion alignment & missing data	Cross-modal synchronization frameworks	AMIGOS, ASCERTAIN
44	Torres et al. ⁵⁵	XAI methods inconsistent	Reliable explainable AI for EEG	DEAP, SEED
45	Fiorini et al. ⁵⁶	Deep models black-box	Clinically validated interpretability	DEAP, SEED
46	Gkintoni et al. ⁷²	Fragmented evaluation practices	Unified systematic review benchmarks	DEAP, SEED, MAHNOB-HCI
47	Wang et al. ⁷³	Limited focus on temporal dependencies	Temporal transformer integration	DEAP
48	Ganepola et al. ⁷⁴	Narrow emotion taxonomy	Broader affective dimensions	DEAP, SEED
49	Yu et al. ⁵⁴	Transformer benchmark limited to labs	Multi-center testing for robustness	SEED, MAHNOB-HCI

Deep Learning (CNN, LSTM, Transformer): Better performance (~85–95% in CS) and low interpretability.

One such emergent yet promising direction is the combination of Explainable AI (XAI) techniques, such as those used in work by Azar¹ with fuzzy rules. This area however according to the Table 3 (Limitations) requires even more development and standardisation to become clinically meaningful.

Abbreviations and Definitions

Validation Types: SD (Subject-Dependent), CS (Cross-Subject, e.g., LOSO, k-fold across subjects), CSS (Cross-Session), CD (Cross-Dataset)

Emotion Model: Dimensional (V/A = Valence/Arousal), Discrete (e.g., Happy, Sad, Fear, Neutral)

Split Transparency: Full (exact split described), Partial (split type mentioned but lacks detail), None (no description)

Augmentation Timing: Pre-split (applied before train/test split), Post-split (applied only to training data), None

The limitations, problems, and gaps in the research studies of EEG-based emotion recognition have been systematically summarized in Table 3 based on the benchmark databases used, i.e., DEAP, SEED, DREAMER, AMIGOS, ASCERTAIN, and MAHNOB-HCI. The research shows that the least evaded obstacle is the cross-subject and cross-session variability because models tend to generalize outside of the subjects with which the models are trained. Imbalance in the datasets and homogeneity of the demographics also restricts the strength of the models, and most benchmark datasets have limited diversity of participants, recording conditions and ecological validity. Although deep learning methods are highly accurate, they have limitations including high cost of computation, interpretability and reproducibility particularly in real-time and low-data conditions. Besides, difference in feature extraction, preprocessing pipelines and validation protocols do not make cross-study comparisons to be fair. All these observations together point to the urgent need to develop standardized evaluation models, multimodal and demographically diverse data, and lightweight and explainable models that can be generalized across individuals and settings. Trying to classify the existing mass of evidence in terms of the limitations of the datasets used, Table 3 provides the important insight into the current changes in the research field and specifies

the directions of future research that should be taken in order to produce the robust, transparent, and deployable systems of emotion recognition via EEG.

Risk of Bias Assessment

To evaluate the quality and reproducibility of the methodological aspects of the included studies critically, formal risk-of-bias assessment was performed in Table 4, according to a pre-defined rubric. The evaluation targeted four major areas which are important to validate and replicate machine learning research:

Split Transparency: Did the data splitting process (e.g., subject dependent, cross-subject, LOSO) receive sufficient description, with a verbatim description of the composition of the training and test sets?

Data Augmentation & Leakage Protection: Was data augmentation disclosed? Were other leakage safeguards (such as subject-wise normalization) described? Did they use it, did they apply it after training-test split (to avoid leakage) and did they note it?

Validation Integrity: Was the study based on a rigorous validation scheme that could be applied to the real world (e.g., cross-subject or cross-session across subject-dependent) and performance was reported with measures of variance (e.g., standard deviation)?

Openness & Reproducibility: Did the model and evaluation code exist in a publicly accessible place? Did it provide the splits or trained models of data?

Rubric and Scoring

Each domain was rated in relation to each study as follows:

Low risk: The criterion was fully and clearly reported in the study.

Medium Risk: The criterion was partially reported in the study, or it failed to be clear.

High Risk: This study did not state the criterion, or the procedure in which the study was performed placed an evident risk of bias (e.g., augmentation used before splitting).

All included studies have the results of this assessment summarized in Figure 4. This traffic-light chart gives a summary of how the biases have been distributed in the literature.

The figure illustrates the percentage of researches considered as low, medium, or high risk of each bias domain. The stepwise analysis per-study analysis can be found in the supplementary materials.

Domain	Low Risk	Medium Risk	High Risk
Split Transparency	Exact split described (e.g., "LOSO with 32 subjects," "70-15-15 split per subject").	Split type mentioned but lacks detail (e.g., "cross-validation" without specifying k).	No description of how data was split for training/testing.
Augmentation & Leakage Safeguards	Augmentation disclosed and applied post-split ; OR no augmentation used and other safeguards (e.g., subject-wise normalization) stated.	Augmentation disclosed but timing unclear; OR no augmentation and no mention of safeguards.	Augmentation used but timing suggests pre-split (high leakage risk); OR augmentation not disclosed but likely used.
Validation Integrity	Cross-subject/session validation used AND performance variance (SD/CI) reported.	Cross-subject/session validation used BUT no variance reported; OR subject-dependent with variance.	Subject-dependent validation AND no variance reported.
Openness & Reproducibility	Code and data splits or model weights available in a public repository.	Code available but no data splits/models; OR only a non-executable algorithm description.	No code or supplementary materials provided.

Code and Model Availability

Lastly, the reproducibility is constrained by the nature of open-source code and pretrained models which are very limited. Less than two out of ten studies that are reviewed publish their implementation or evaluation scripts. The unavailability of codes hinders the process of independent verification and benchmarking and is also a contributing factor to publication bias where only successful experiments are published. Efforts to popularize open repositories of EEG data, pre-processing software and trained models, including the public benchmark portal by SEED, should be expanded to all large emotion datasets.

Risk of Bias Assessment Rubric

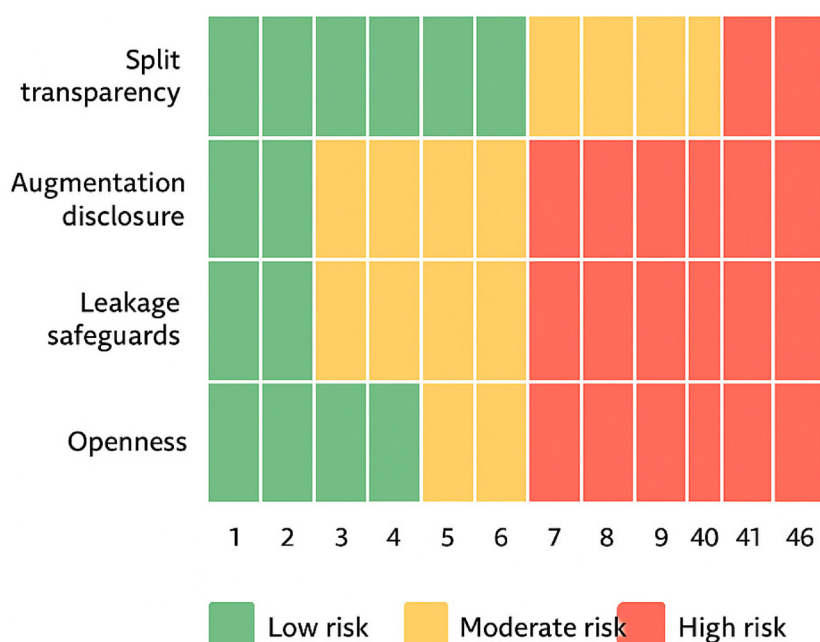


Fig 4 | Summary of risk assessment of bias in all the studies included

Category	Checklist Items for Practitioners
Privacy & Consent	<ul style="list-style-type: none"> Obtain explicit, revocable consent. Encrypt and anonymize raw EEG data. Document data retention and reuse policies.
Fairness & Inclusion	<ul style="list-style-type: none"> Report demographics of participants. Test model fairness across subgroups. Use balanced or stratified datasets.
Transparency & Reproducibility	<ul style="list-style-type: none"> Release preprocessing and training code. Publish split definitions and random seeds. Share trained model weights (when permissible).
Calibration & Stability	<ul style="list-style-type: none"> Minimize calibration time per user. Validate model performance across sessions/devices. Include long-term drift analysis.
Latency & Resource Budgets	<ul style="list-style-type: none"> Report inference latency. Quantify memory and compute requirements. Optimize models for edge or embedded systems.
Ethical Oversight	<ul style="list-style-type: none"> Obtain IRB or ethics committee approval. Provide user opt-out and system transparency. Ensure emotion feedback is non-invasive and non-manipulative.

Deployment and Ethical Considerations

Privacy and Data Governance

EEG signals are distinctively identifiable and capable of displaying emotional states as well as health and cognitive data, which makes privacy protection the main priority. The principles of data minimization, limit purpose, and informed consent must all be applied to all data processing.

Anonymization can be too little as EEG patterns can be re-identified between sessions and datasets. Thus, cross-institutional training should be taken into account using privacy-preserving learning, including federated learning, differential privacy, or secure multi-party computation. Besides, GDPR and local bioethics require dataset custodians to expressly specify storage periods, encryption norms, and access rights of users.

Fairness and Demographic Imbalance

The existing EEG-based emotion datasets (e.g., DEAP, SEED, DREAMER) have biased demographics (an overall overrepresentation of young, male, university-educated participants of limited ethnic background). Such an unequal distribution creates the danger of introducing prejudice into the classifiers resulting in unequal performance between genders, ages, or cultural backgrounds. Subsequent datasets ought to embrace stratified sampling and demographic balancing measures and published models ought to incorporate subgroup performance measurements. The researchers ought to not only provide the composition of the dataset but also provide the possible bias during the electrode placement, interpretation of emotional stimuli, or language-specific labeling of affect.

Informed Consent and Participant Autonomy

EEG ethical studies should assure that the participants are aware of:

1. The type and the length of EEG data recording.
2. The utilized emotional stimuli (and possible psychological influence).
3. Policies of future reuse and sharing.

The consent procedures must be continuous not single especially in longitudinal studies. Anytime models are put out in either a social or a clinical context, users should have the right to switch off emotion monitoring, and the system will need to show the status (e.g. when recording is on).

Calibration and User Burden

Normally EEG emotion systems must be calibrated on a per-user basis to normalize features. Although calibration provides more accuracy, it adds more burden to the end-users. Research to minimize this dependency is currently moving towards cross-subject generalization and transfer learning methods that allow plug and play emotion recognition with minimal retraining. Nevertheless, even the calibration-free models are expected to be tested to have long-term stability, session

drift, and hardware variability. The accuracy decay may be reduced without a lot of effort by regular recalibration schedules (e.g., quarterly).

Real-Time Constraints and Resource Budgets

Real-time inference in severe latency and memory constraints is needed in applications such as wearables, robotics, or human to computer interaction. In a real-time EEG pipeline, standard latency requirements are less than 150 ms since this is responsive to adaptive feedback systems.

Memory and compute budgets should align with embedded systems

- Mobile or edge processing should use less than 500MB RAM, less than 1W of energy.
- On-device applications have working models pruned at train, quantized, and lightweight (MobileNet, TinyCNN, SpikingNN).

Computational footprints and latency benchmarks must be indicated together with accuracy to enable a transparent trade-off between speed and performance as indicated in Table 5 practitioner checklist for responsible EEG-emotion pipelines.

Open Science, Reproducibility and Transparency

The privacy and fairness are not the only ethical issues of ethical deployment that reach into the scientific reproducibility. Whenever feasible, all code, preprocessing scripts and model weights trained should be made publicly available under open licenses (e.g. MIT, CC-BY). Researchers are required to record:

EEG preprocessing pipelines (filtering, artifact removal).

- Strategies of feature extraction and normalization.
- Definitions and random seeds of training/tests.
- Open information sharing minimizes redundancy and enhances community validation, as well as securing intellectual property and privacy of the participants.

Conclusion

A comprehensive review of selected studies reveals substantial advancements in EEG-based emotion recognition, driven by both traditional and deep learning approaches. The profound impact of emotions on human behaviour and decision-making, the accurate detection and interpretation of emotional states hold substantial application value across healthcare, education, and entertainment. With the advancement of brain-computer interface (BCI) technologies and artificial intelligence, EEG-based emotion recognition has gained significant momentum in recent years. This review has outlined the critical processes involved in EEG-based emotion recognition. It emphasized that the processes of signal acquisition and pre-processing play a crucial role in determining the accuracy of emotion classification. Furthermore, the choice of classification method significantly impacts the reliability of recognition results. With

the successful application of deep learning techniques in this field, researchers have proposed a variety of neural network-based models. In particular, hybrid neural network architectures that combine different deep learning models have shown strong potential in capturing complex EEG patterns. These models, particularly when integrated with topographic feature maps and connectivity matrices, excel in capturing spatial-temporal patterns in EEG data. Ensemble techniques, including rotation forests, further enhance robustness. Overall, the reviewed literature confirms the continued efforts in this area are expected to further enhance the accuracy, robustness, and real-world applicability of emotion recognition technologies.

Future Work

The architecture transformer has shown itself to be better performing in many fields recently because they are capable of capturing long-range dependencies and more complex spatiotemporal relationships. Transformers in EEG emotion recognition work by providing the ability to record inter-channel correlations and time behavior in parallel without having to employ recurrent structures. Other models like Multi-Scale Dual Channel Graph Transformer Network (MSDCGTNet) combine attention mechanisms and graph-based representations to learn patterns of spatial connectivity between different regions of the brain. Transformer architectures are however computationally intensive, which means that they need large labeled datasets and large training resources. The research on transformers should thus focus on effective versions of transformers to include lightweight or hybrid CNN-transformer models that retain accuracy and allow real-time processing. Attention mechanisms that take into consideration neurophysiological priors can also be more interpretable in terms of mapping a learned attention map to familiar emotional circuitry.

The second direction is emerging as self-supervised EEG representation learning, which attempts to alleviate a situation where labeled data is limited, a significant bottleneck in the area. The classic supervised processes are based on small datasets that are manually labeled and thus restrict generalization. Self-supervised learning (SSL) enables models to train on unsupervised large data sets of intrinsic EEG to learn intrinsic EEG representations via contrastive or masked signal reconstruction tasks. SSL frameworks can be unsupervisedly pre-trained on limited emotion-labeled samples, and even in few-shot scenarios, fine-tuning them can be done with high performance. This method does not only provide efficiency in data, but also decreases the reliance on certain datasets, which leads to more generalizable and transferable feature representations. The benchmarking of various strategies of the SSL such as those based on temporal contrastive learning and masked autoencoders into the most appropriate formulations should be conducted in future works to apply them in the non-stationary nature of EEG.

The concept of real-time emotion decoding is another important research area of interest that seeks to

apply laboratory models in real practice. There is a limitation in most of the existing systems; they are only tested offline, which restricts their use in affective computing, adaptive learning and healthcare monitoring. The use of decoding emotion in real-time needs minimal-weight architectures that can primarily implement continuous inference at low latency and power consumption. Model pruning, knowledge distillation, and on-device quantization are the strategies that can help to dramatically decrease the amount of computations without accuracy loss. Moreover, by combining streaming EEG pipelines with edge devices or wearable devices, it would be easier to deploy emotion-aware systems in a more naturalistic setting. The studies in this field should also cover the latency-compensation methods, and the latency adjustment to signal drift with time.

Finally, cross-device and cross-dataset generalization is also an unresolved problem having a direct impact on the robustness and reproducibility of models. Variations in EEG devices, electrode arrangements or recording conditions can cause domain shifts which reduce performance when models are moved between devices or subject groups. Research in the future should come up with domain adaptation models with the ability to match representations between heterogeneous EEG sources. This can be adversarial learning, subspace alignment or meta-learning mechanisms which encourage invariance to device-specific noise. It will be necessary to have an open cross-device benchmark, and a set of standardized preprocessing pipelines to be able to fairly compare and reproducibly.

References

- Mert A, Akan A. Emotion recognition based on time–frequency distribution of EEG signals using multivariate synchrosqueezing transform. *Digit Signal Process.* 2018;81:106–15. <https://doi.org/10.1016/j.dsp.2018.07.003>
- Moon SE, Chen CJ, Hsieh CJ, Wang JL, Lee JS. Emotional EEG classification using connectivity features and convolutional neural networks. *Neural Netw.* 2020;132:96–107. <https://doi.org/10.1016/j.neunet.2020.08.009>
- Topic A, Russo M. Emotion recognition based on EEG feature maps through deep learning network. *Eng Sci Technol Int J.* 2021;24(6):1442–54. <https://doi.org/10.1016/j.jestch.2021.03.012>
- Liu ZT, Xie Q, Wu M, Cao WH, Li DY, Li SH. Electroencephalogram emotion recognition based on empirical mode decomposition and optimal feature selection. *IEEE Trans Cogn Dev Syst.* 2018;11(4):517–26. <https://doi.org/10.1109/TCDS.2018.2878696>
- Kuang F, Shu L, Hua H, Wu S, Zhang L, Xu X. Cross-subject And Cross-device Wearable EEG Emotion Recognition Using Frontal EEG Under Virtual Reality Scenes. In: 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 2021. p. 3630–7. <https://doi.org/10.1109/BIBM52615.2021.9669802>
- Azar NAN, Cavus N, Esmaili P, Sekeroglu B, Aşır S. Detecting Emotions Through EEG Signals Based on Modified Convolutional Fuzzy Neural Network. *Sci Rep.* 2024;14:10371. <https://doi.org/10.1038/s41598-024-60977-9>
- Khan A, Hussain M, Anwar H, Khan MU. Developing an EEG-based emotion recognition system using machine learning. *IEEE Access.* 2023;11:1869–83. <https://doi.org/10.1109/ACCESS.2023.3230001>
- Rahman M, Hasan MT, Al-Qaysi AM, Zahid MAH. Emotion detection from EEG signals using machine and deep learning: a comparative study. *Sensors.* 2022;22(17):6550. <https://doi.org/10.3390/s22176550>
- Chen H, Zhang Y, Liu Y. Emotion recognition from EEG signals using recurrent neural networks with attention mechanism. *IEEE Access.* 2021;9:19656–66. <https://doi.org/10.1109/ACCESS.2021.3053467>
- Wang Y, Lu S, Zhang L. Human emotion recognition from EEG-based brain-computer interface using hybrid deep neural network. *IEEE Trans Cogn Dev Syst.* 2021;13(2):354–64. <https://doi.org/10.1109/TCDS.2020.2992063>
- Liu F, Liu G, Wang H. Strengthen EEG-based emotion recognition using firefly integrated metaheuristic learning. *Inf Fusion.* 2021;67:57–68. <https://doi.org/10.1016/j.inffus.2020.10.004>
- Singh R, Sharma VK. Multi-channel EEG-based emotion recognition via a multi-level features fusion approach. *Biocybern Biomed Eng.* 2020;40(4):1496–508. <https://doi.org/10.1016/j.bbe.2020.08.003>
- Li B, Liu Y, Li J. Emotion recognition with machine learning using EEG signals: a review. *Biomed Signal Process Control.* 2020;58:101838. <https://doi.org/10.1016/j.bspc.2020.101838>
- Patel D, Chauhan R. Emotions recognition using EEG signals: a comprehensive review. *Mater Today Proc.* 2023;72:2677–82. <https://doi.org/10.1016/j.matpr.2023.02.104>
- Alarcao S, Fonseca MJ. EEG-based emotion recognition: a tutorial and review. *ACM Comput Surv.* 2019;51(6):1–36. <https://doi.org/10.1145/3277668>
- Bagherzadeh S, Shalhaf A, Shoeibi A, Jafari M, Tan RS, Acharya UR. Developing an EEG-Based Emotion Recognition Using Ensemble Deep Learning Methods and Fusion of Brain Effective Connectivity Maps. *IEEE Access.* 2023;12:50949–65. <https://doi.org/10.1109/ACCESS.2024.3384303>
- Fu B, Li F, Niu Y, Wu H, Li Y, Shi G. Conditional generative adversarial network for EEG-based emotion fine-grained estimation and visualization. *J Vis Commun Image Represent.* 2021;74:102982. <https://doi.org/10.1016/j.jvcir.2020.102982>
- Liu Y, Fu G. Emotion recognition by deeply learned multi-channel textual and EEG features. *Future Gener Comput Syst.* 2021;119:1–6. <https://doi.org/10.1016/j.future.2021.01.010>
- Gong L, Li M, Zhang T, Chen W. EEG emotion recognition using attention-based convolutional transformer neural network. *Biomed Signal Process Control.* 2023;84:104835. <https://doi.org/10.1016/j.bspc.2023.104835>
- Liu Y, Ding Y, Li C, Cheng J, Song R, Wan F, et al. Multi-channel EEG-based emotion recognition via a multi-level features guided capsule network. *Comput Biol Med.* 2020;123:103927. <https://doi.org/10.1016/j.cbi.2020.103927>
- He H, Tan Y, Ying J, Zhang W. Strengthen EEG-based emotion recognition using firefly integrated optimization algorithm. *Appl Soft Comput.* 2020;94:106426. <https://doi.org/10.1016/j.asoc.2020.106426>
- Gao Z, Li Y, Yang Y, Wang X, Dong N, Chiang HD. A GPSO-optimized convolutional neural networks for EEG-based emotion recognition. *Neurocomputing.* 2020;380:225–35. <https://doi.org/10.1016/j.neucom.2019.10.096>
- Cui H, Liu A, Zhang X, Chen X, Wang K, Chen X. EEG-based emotion recognition using an end-to-end regional-asymmetric convolutional neural network. *Knowl Based Syst.* 2020;205:106243. <https://doi.org/10.1016/j.knsys.2020.106243>
- Subasi A, Tuncer T, Dogan S, Tanko D, Sakoglu U. EEG-based emotion recognition using tunable Q wavelet transform and rotation forest ensemble classifier. *Biomed Signal Process Control.* 2021;68:102648. <https://doi.org/10.1016/j.bspc.2021.102648>
- Islam MR, Islam MM, Rahman MM, Mondal C, Singha SK, Ahmad M, et al. EEG Channel Correlation Based Model for Emotion Recognition. *Comput Biol Med.* 2021;136:104757. <https://doi.org/10.1016/j.combiomed.2021.104757>
- Atkinson J, Campos D. Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers. *Expert Syst Appl.* 2016;47:35–41. <https://doi.org/10.1016/j.eswa.2015.10.049>
- Lu W, Liu H, Ma H, Tan TP, Xia L. Hybrid transfer learning strategy for cross-subject EEG emotion recognition. *Front Hum Neurosci.* 2023;17:1280241. <https://doi.org/10.3389/fnhum.2023.1280241>
- Jiménez-Guarneros M, Fuentes-Pineda G. Learning a Robust Unified Domain Adaptation Framework for Cross-Subject EEG-Based Emotion Recognition. *Biomed Signal Process Control.* 2023;86:105138. <https://doi.org/10.1016/j.bspc.2023.105138>
- Luo T, Zhang J, Qiu Y, Zhang L, Hu Y, Yu Z, et al. M3D: Manifold-based Domain Adaptation with Dynamic Distribution for Non-Deep

- Transfer Learning in Cross-subject and Cross-session EEG-based Emotion Recognition. *IEEE J Biomed Health Inform.* 2025;1–21. <https://doi.org/10.1109/JBHI.2025.3580612>
- 30 Li J, Hua H, Xu Z, Shu L, Xu X, Kuang F, et al. Cross-subject EEG emotion recognition combined with connectivity features and meta-transfer learning. *Comput Biol Med.* 2022;145:105519. <https://doi.org/10.1016/j.compbiomed.2022.105519>
- 31 Zheng WL, Lu BL. Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks. *IEEE Trans Auton Ment Dev.* 2015;7(3):162–75. <https://doi.org/10.1109/TAMD.2015.2431497>
- 32 Chen J, Jiang D, Zhang Y, Zhang P. Emotion recognition from spatiotemporal EEG representations with hybrid convolutional recurrent neural networks via wearable multi-channel headset. *Comput Commun.* 2020;154:58–65. <https://doi.org/10.1016/j.comcom.2020.02.051>
- 33 Akhand MAH, Maria MA, Kamal MAS. Improved EEG-based emotion recognition through information enhancement in connectivity feature map. *Sci Rep.* 2023;13:13804. <https://doi.org/10.1038/s41598-023-40786-2>
- 34 Chowdary MK, Anitha J, Hemanth DJ. Emotion Recognition from EEG Signals Using Recurrent Neural Networks. *Electronics.* 2022;11(15):2387. <https://doi.org/10.3390/electronics11152387>
- 35 Zhang Z, Lu G. Multimodal Knowledge Distillation for Emotion Recognition. *Brain Sci.* 2024;15(7):707. <https://doi.org/10.3390/brainsci15070707>
- 36 Cheng Z, Bu X, Wang Q, et al. EEG-based emotion recognition using multi-scale dynamic CNN and gated transformer. *Sci Rep.* 2024;14:31319. <https://doi.org/10.1038/s41598-024-82705-z>
- 37 Liu Q, Hao J, Guo Y. EEG Data Augmentation for Emotion Recognition with a Task-Driven GAN. *Algorithms.* 2023;16(2):118. <https://doi.org/10.3390/a16020118>
- 38 Song Y, Feng L, Zhang W, Song X, Cheng M. Multimodal Emotion Recognition based on the Fusion of EEG Signals and Eye Movement Data. In: 2024 IEEE 25th China Conference on System Simulation Technology and its Application (CCSSTA); 2024. p. 127–32. <https://doi.org/10.1109/CCSSTA62096.2024.10691734>
- 39 Wang F, Tian YC, Zhou X. Cross-dataset EEG emotion recognition based on pre-trained Vision Transformer considering emotional sensitivity diversity. *Expert Syst Appl.* 2025;279:127348. <https://doi.org/10.1016/j.eswa.2025.127348>
- 40 Imtiaz MN, Khan N. Enhanced cross-dataset electroencephalogram-based emotion recognition using unsupervised domain adaptation. *Comput Biol Med.* 2025;184:109394. <https://doi.org/10.1016/j.compbiomed.2024.109394>
- 41 Khan SA, Chaudary E, Mumtaz W. EEG-ConvNet: Convolutional networks for EEG-based subject-dependent emotion recognition. *Comput Electr Eng.* 2024;116:109178. <https://doi.org/10.1016/j.compeleeng.2024.109178>
- 42 Alghamdi AM, Ashraf MU, Bahaddad AA, et al. Cross-subject EEG signals-based emotion recognition using contrastive learning. *Sci Rep.* 2025;15:28295. <https://doi.org/10.1038/s41598-025-13289-5>
- 43 Alameer HRA, Salehpour P, Aghdasi HS, Feizi-Derakhshi MR. Integrating Deep Metric Learning, Semi Supervised Learning, and Domain Adaptation for Cross-Dataset EEG-Based Emotion Recognition. *IEEE Access.* 2025;13: 38914–24. <https://doi.org/10.1109/ACCESS.2025.3536549>
- 44 Patel P, Balasubramanian S, Annavarapu RN. Cross subject emotion identification from multichannel EEG sub-bands using Tsallis entropy feature and KNN classifier. *Brain Inf.* 2024;11(7):1–13. <https://doi.org/10.1186/s40708-024-00220-3>
- 45 Rakhmatulin I, Dao M-S, Nassibi A, Mandic D. Exploring Convolutional Neural Network Architectures for EEG Feature Extraction. *Sensors.* 2024;24(3):877. <https://doi.org/10.3390/s24030877>
- 46 Feng S, Wu Q, Zhang K, Song Y. A Transformer-Based Multimodal Fusion Network for Emotion Recognition Using EEG and Facial Expressions in Hearing-Impaired Subjects. *Sensors.* 2025;25(20):6278. <https://doi.org/10.3390/s25206278>
- 47 Tan W, Zhang H, Wang Y, Wen W, Chen L, Li H, et al. SEDA-EEG: A semi-supervised emotion recognition network with domain adaptation for cross-subject EEG analysis. *Neurocomputing.* 2025;622:129315. <https://doi.org/10.1016/j.neucom.2024.129315>
- 48 An Y, Lam HK, Ling SH. Multi-classification for EEG motor imagery signals using data evaluation-based auto-selected regularized FBSP and convolutional neural network. *Neural Comput Appl.* 2023;35:12001–27. <https://doi.org/10.1007/s00521-023-08336-z>
- 49 Manoj Prasath T, Vasuki R. Integrated Approach for Enhanced EEG-Based Emotion Recognition with Hybrid Deep Neural Network and Optimized Feature Selection. *Int J Electron Commun Eng.* 2023;10(11):55–68. <https://doi.org/10.14445/23488549/IJECE-V10I11P106>
- 50 Soleymani M, Lichtenauer J, Pun T, Pantic M. A multimodal database for affect recognition and implicit tagging. *IEEE Trans Affect Comput.* 2012; 3(1): 42–55. <https://doi.org/10.1109/T-AFFC.2011.25>
- 51 Subramanian R, Wache J, Abadi MK, Vieriu R, Winkler S, Sebe N. ASCERTAIN: Emotion and Personality Recognition Using Commercial Sensors. *IEEE Trans Affect Comput.* 2018;9(2):147–60. <https://doi.org/10.1109/TAFFC.2016.2625250>
- 52 Katsigiannis S, Ramzan N. DREAMER: A Database for Emotion Recognition Through EEG and ECG Signals from Wireless Low-Cost Off-the-Shelf Devices. *IEEE J Biomed Health Inform.* 2017;22(1):98–107. <https://doi.org/10.1109/JBHI.2017.2688239>
- 53 Miranda-Correa JA, Abadi MK, Sebe N, Patras I. AMIGOS: A Dataset for Affect, Personality and Mood Research on Individuals and Groups. *IEEE Trans Affect Comput.* 2021;12(2):479–93. <https://doi.org/10.1109/TAFFC.2018.2884461>
- 54 Yu L, Ge Y, Ansari S, Imran M, Ahmad W. Multimodal sensing-enabled large language models for automated emotional regulation: a review of current technologies, opportunities, and challenges. *Sensors.* 2025;25(15):4763. <https://doi.org/10.3390/s25154763>
- 55 Mayor Torres JM, Medina-DeVilliers S, Clarkson T, Lerner MD, Riccardi G. Evaluation of interpretability for deep learning algorithms in EEG emotion recognition: a case study in autism. *Artif Intell Med.* 2023;143:102545. <https://doi.org/10.1016/j.artmed.2023.102545>
- 56 Fiorini L, Bossi F, Di Gruttola F. EEG-based emotional valence and emotion regulation classification: a data-centric and explainable approach. *Sci Rep.* 2024;14:24046. <https://doi.org/10.1038/s41598-024-75263-x>
- 57 Liu R, Chao Y, Ma X, Sha X, Sun L, Li S, Chang S. ERTNet: an interpretable transformer-based framework for EEG emotion recognition. *Front Neurosci.* 2024;18:1320645. <https://doi.org/10.3389/fnins.2024.1320645>
- 58 Koelstra S, Mühl C, Soleymani M, Lee JS, Yazdani A, Ebrahimi T, DEAP: A Database for Emotion Analysis Using Physiological Signals. *IEEE Trans Affect Comput.* 2012;3(1):18–31. <https://doi.org/10.1109/T-AFFC.2011.15>
- 59 Yuvaraj R, Baranwal A, Prince AA, Murugappan M, Mohammed JS. Emotion recognition from spatio temporal representation of EEG signals via 3D CNN with ensemble learning techniques. *Brain Sci.* 2023;13(4):685. <https://doi.org/10.3390/brainsci13040685>
- 60 Yu X, Li Z, Zhang Z, Liu Y. Real-time EEG-based emotion recognition. *Sensors.* 2023;23(18):7853. <https://doi.org/10.3390/s23187853>
- 61 Zhang M, Yang J, Liu Y, Zhang X. Detecting emotions through EEG signals based on modified convolutional fuzzy neural network. *IEEE Trans Fuzzy Syst.* 2022;30(8):3233–43. <https://doi.org/10.1109/TFUZZ.2021.3098332>
- 62 Hamzah MA, Abdalla A. EEG-based emotion recognition systems: a comprehensive study. *Multimed Tools Appl.* 2024;83:1825–64. <https://doi.org/10.1007/s11042-023-15507-4>
- 63 Ma J, Yang B, Qiu W, Li Y, Zhao N, He H. A large EEG dataset for studying cross session variability in motor imagery brain computer interface. *Sci Data.* 2022;9(1):531. <https://doi.org/10.1038/s41597-022-01647-1>
- 64 Yin Y, Wang P, Childs PRN. Understanding creativity process through electroencephalography measurement on creativity related cognitive factors. *Front Neurosci.* 2022;16:951272. <https://doi.org/10.3389/fnins.2022.951272>
- 65 Dhara T, Singh PK, Mahmud M. A fuzzy ensemble based deep learning model for EEG based emotion recognition. *Cogn Comput.* 2024;16:1364–78. <https://doi.org/10.1007/s12559-023-10171-2>
- 66 Jirayucharoensak S, Pan-Num S, Israsena P. EEG-based emotion recognition using deep learning network with principal component

- based covariate shift adaptation. *Sci World J.* 2014;2014:627892. <https://doi.org/10.1155/2014/627892>
- 67 Liu X, Wang B, Wang J, Wang S, Yan J, Teng Q, You W. Effect of transcutaneous acupoint electrical stimulation on propofol sedation: an electroencephalogram analysis of patients undergoing pituitary adenomas resection. *BMC Complement Altern Med.* 2016;16(1):33. <https://doi.org/10.1186/s12906-016-1008-1>
- 68 Wang YT, Huang KC, Wei CS, Huang TY, Ko LW, Lin CT, Cheng CK, Jung TP. Developing an EEG based on-line closed loop lapse detection and mitigation system. *Front Neurosci.* 2014;8:321. <https://doi.org/10.3389/fnins.2014.00321>
- 69 Zheng WL, Zhu JY, Peng Y, Lu BL. EEG-Based Emotion Classification Using Deep Belief Networks. In: 2014 IEEE International Conference on Multimedia and Expo (ICME); 2014. p. 1–6. <https://doi.org/10.1109/ICME.2014.6890166>
- 70 Zheng W, Liu W, Lu Y, Lu B, Cichocki A. Emotion Meter: A multimodal framework for recognizing human emotions. *IEEE Trans Cybern.* 2019; 49(3):1110–22. <https://doi.org/10.1109/TCYB.2018.2797176>
- 71 Pillalamarri R, Shanmugam U. A review on EEG based multimodal learning for emotion recognition. *Artif Intell Rev.* 2025;58(5):131. <https://doi.org/10.1007/s10462-025-11126-9>
- 72 Gkintoni E, Aroutzidis A, Antonopoulou H, Halkiopoulos C. From neural networks to emotional networks: a systematic review of EEG based emotion recognition in cognitive neuroscience and real world applications. *Brain Sci.* 2025;15(3):220. <https://doi.org/10.3390/brainsci15030220>
- 73 Wang W, Huang M, Wang R, Zhang L. Deep learning-based EEG emotion recognition: current trends and future perspectives. *Front Neurosci.* 2020;14:570746. <https://doi.org/10.3389/fnins.2020.570746>
- 74 Ganepola D, Maduranga MWP, Tilwari V, Karunaratne I. A systematic review of electroencephalography based emotion recognition of confusion using artificial intelligence. *Signals.* 2024;5(2):244–63. <https://doi.org/10.3390/signals5020013>
- 75 Wu R. Analysis of emotion recognition based on brain-computer interface technology. *Theor Natl Sci.* 2023;18:281–9. <https://doi.org/10.54254/2753-8818/18/20230443>