Preliminary Exploration of Nine-Class EEG Emotion Recognition with the FACED Dataset

Aaron Tan (Mentee)

Caitrin Burke (Mentor)

aarntn1232@gmail.com

Youth Neuropsychology Society

Abstract

Decoding emotion from scalp EEG remains challenging at scale, as nine-class, cross-subject emotion recognition from scalp EEG remains under-examined; most studies use two or three classes and intra-subject validation, which inflates performance. Hence, this study presents a strict, reproducible pipeline on the FACED dataset (32 channels, 123 participants, 28 clips). Preprocessing applies 0.05–47 Hz band-pass filtering, downsampling to 250 Hz, ICA-based artifact reduction using EEG-only criteria, re-referencing, and segmentation into 1-s epochs. Features combine differential entropy and power spectral density across five bands (1–4, 4–8, 8–14, 14–30, 30–47 Hz), aggregated into trial-level vectors and standardized using training-partition statistics only. A compact multilayer perceptron (two 256-unit ReLU layers with dropout, Adam) is trained with 10-fold StratifiedGroupKFold by subject; a subject-wise validation split controls early stopping. All transforms, ICA, and interpolation are fitted on training subjects within each fold and then applied to validation and testing to prevent leakage. Evaluation reports macro-F1, balanced accuracy, MCC, per-class precision/recall, and PR/ROC AUC with means and 95% confidence intervals across folds, alongside probability calibration (ECE/MCE), class-imbalance handling, and baselines (logistic regression, SVM, and random forest). Moreover, this study includes ablations (DE-only, PSD-only), sensitivity to band edges and epoch length, and cross-cohort transfer tests. The protocol emphasizes transparency over headline scores and offers a reference protocol for nine-class EEG emotion recognition on FACED. Code, fold indices, and figure scripts accompany the study.

Introduction

Emotions shape cognition, decision-making, and behavior. Yet most EEG-based emotion recognition studies still use two or three labels and intra-subject evaluation, which overestimates performance and limits real-world use (Liu et al., 2021; Brunner-Sperdin et al., 2012; Ramirez et al., 2010; Xin et al., 2019; Zhang et al., 2024). The field needs fine-grained, multi-class recognition that generalizes across unseen people.

Current public corpora are small and coarsely labeled. DEAP (32 participants; 40 music videos) targets valence/arousal and is often binarized (Koelstra et al., 2011). SEED (15 participants; film clips) uses three categories: positive, neutral, negative (BCMI, 2015). DREAMER (23 participants; film clips; EEG+ECG) also centers on valence/arousal (Katsigiannis & Ramzan, 2017). These datasets rarely distinguish discrete positive states such as joy, amusement, or inspiration. Modest sample sizes and limited diversity further constrain cross-subject claims. For practical deployment, models must be trained on some people and tested on different people (Miranda-Correa et al., 2018).

Comparability across studies is weak. Feature sets, preprocessing, and metrics differ; systematic model comparisons under identical conditions are uncommon; and code or processed data are often unavailable (Cohen et al., 2023; Ahmed et al., 2023). Standardized, leakage-free protocols with full reporting are required to assess genuine progress.

EEG offers millisecond-scale temporal resolution and supports analysis in canonical bands: delta (1–4 Hz), theta (4–8 Hz), alpha (8–14 Hz), beta (14–30 Hz), and a conservative "gamma" (30–47 Hz). Alpha suppression (event-related desynchronization) indexes increased cortical engagement; midline frontal theta often tracks working memory and affective control. We exclude frequencies >47 Hz and treat 30–47 Hz cautiously because

scalp “gamma” can include EMG; interpretations focus on spectral descriptors rather than strong cognitive claims (Singh et al., 2025; Al-Hadithy et al., 2025; Woodman et al., 2021).

To represent these rhythms, we use differential entropy (DE) and power spectral density (PSD). Under a Gaussian assumption, DE reduces to log-variance and summarizes band-specific variability (Shi et al., 2013). PSD estimates band-limited power linked to emotion–cognition interactions (Wang et al., 2014). The two feature families are complementary—PSD captures amplitude structure, and DE captures dispersion—and form the basis of our analysis.

The FACED dataset (Chen et al., 2023) enables fine-grained, subject-independent evaluation. It provides 32-channel EEG from 123 participants who viewed 28 curated clips eliciting nine discrete states: joy, tenderness, inspiration, amusement, anger, fear, disgust, sadness, and neutral. Alongside EEG, FACED includes self-reports (valence, arousal, liking, familiarity) and a cross-subject design. We adopt strict, leakage-free processing and evaluation on FACED to quantify nine-class performance, report uncertainty across folds, and compare models under matched conditions.

Hence, we use FACED to establish a reference protocol for nine-class, cross-subject EEG emotion recognition. The study asks: (i) how well do multilayer perceptrons trained on DE+PSD features classify fine-grained emotions relative to strong linear and tree baselines, and (ii) what structure is visible in the learned representations via dimensionality reduction and scalp maps?

We hypothesize that an MLP will outperform linear baselines by modeling non-linear interactions among band-limited features, and that confusions will reflect psychological proximity in valence/arousal. Evaluation uses 10-fold subject-wise StratifiedGroupKFold

with all transforms fit on training subjects only. Primary endpoints are macro-F1 and MCC; secondary endpoints include balanced accuracy, macro/micro AUC, and calibration error.

For interpretability, we project penultimate-layer activations with UMAP and quantify scalp patterns using permutation-thresholded topographies of band-limited DE. We also report class-imbalance handling, cohort harmonization, and sensitivity to band edges and epoch length.

This work contributes a clear, reproducible benchmark on FACED: a fixed pipeline, leakage-free splits, baselines under matched conditions, uncertainty reported across folds, and code with fold indices and plotting scripts. The results quantify achievable nine-class performance with DE+PSD features and provide constrained, testable insights into error structure and spatial–spectral patterns.

Methodology

Research Design

This study adopts a computational experimental research design centered on in silico analysis of electroencephalographic (EEG) data for multi-class emotion recognition. Leveraging the large-scale, publicly available FACED dataset, all experimental procedures—from data preprocessing and feature extraction to model training and evaluation—were conducted entirely in silico, without the need for new data collection or wet-lab experimentation. This approach enables systematic benchmarking of machine learning algorithms under controlled, reproducible conditions.

Dataset Description

Table 1. Key Features of the FACED Dataset

Attribute	Summary	Description
Number of subjects	123 participants	EEG data was recorded from a large, demographically diverse sample, enhancing statistical power and cross-subject generalization.
Emotion categories	9 discrete emotions	Includes anger, fear, disgust, sadness, amusement, inspiration, joy, tenderness, and neutral, supporting fine-grained emotion recognition.
Number of video clips	28 curated clips	Each emotion is represented by multiple validated video stimuli (3 per emotion; 4 for neutral), designed to evoke distinct affective responses.
Self-report ratings	12 items on 0–7 scale	Participants rated each video on 8 emotional dimensions and 4 affective attributes.
Recorded signals	32-channel EEG	EEG signals were recorded at high resolution, allowing for frequency-based and entropy-based feature extraction across cortical regions.

We used the Film-based Affective EEG Dataset (FACED) (Chen et al., 2023), which includes 32-channel EEG recordings from 123 participants. Each subject viewed 28 short film clips validated to elicit nine discrete emotions: joy, tenderness, inspiration, amusement, anger, disgust, fear, sadness, and neutral. For every trial, EEG was recorded at 1000 Hz

(cohort 2) or 250 Hz (cohort 1) with synchronized self-report ratings of valence, arousal, liking, and familiarity. Though, it is important to take into account that each of the eight discrete emotion categories has ~369 trials, while Neutral has 492 trials (123 × 28 = 3,444 total). All per-class recalls are reported as percentages (mean ± SD across folds) and not as raw counts. In this study, we analyzed only the EEG signals; all participant recruitment, stimulus presentation, and behavioral data collection were performed by the dataset authors.

Preprocessing

All preprocessing used MNE-Python v1.2.1 (Python 3.10). Continuous EEG was notch-filtered at 50 Hz, then bandpass-filtered at 0.05–47 Hz with an FIR windowed-sinc design. Cohort-2 recordings (1,000 Hz) were resampled to 250 Hz with anti-alias decimation to match cohort-1. Channels were aligned to the 10–20 montage, bad leads were marked, and data were re-referenced to the common average.

Outlier detection operated at the epoch and channel levels. Each 30-s trial was segmented into 30 non-overlapping 1-s epochs. For each recording and channel, we computed the peak-to-peak amplitude distribution; an epoch was flagged if any channel exceeded 3×MAD of that channel's distribution. Channels with >30% flagged epochs in a recording were labeled unreliable and interpolated using spherical splines. Thresholds (3×MAD and 30%) were fixed from the training subjects within each fold and then applied unchanged to validation/test.

Independent component analysis (ICA) was fit per recording on continuous, filtered data (no labels), using FastICA with n_components set to retain 99% variance and a fixed random_state. Components were marked as ocular/motion if they showed frontal topographies plus blink-template correlation >0.30 and/or kurtosis >8; these components were removed. No dedicated EOG channels were available, so identification relied on

EEG-derived criteria. After ICA and interpolation, trials remained segmented into 30×1-s epochs for feature extraction. We log the fraction of retained epochs and the number of interpolated channels per recording.

Feature Extraction

Preprocessed EEG signals were transformed into representations suitable for classification using two complementary feature types. From each 1-s epoch, features were computed across 32 channels and five canonical frequency bands: delta (1–4 Hz), theta (4–8 Hz), alpha (8–14 Hz), beta (14–30 Hz), and gamma (30–47 Hz). Frequencies above 47 Hz were excluded by design to reduce line noise and potential EMG contamination.

Differential entropy (DE, under a Gaussian assumption) was computed per band as $H(x) = \frac{1}{2} log(2\pi e\sigma^2)$ where σ^2 is the variance of the band-pass signal. In practice this is equivalent, up to an affine constant, to $log\sigma^2$, which we applied per channel–epoch–band. This yielded a [32, 30, 5] tensor per trial (channels × epochs × bands), flattened into 4,800 features. Power spectral density (PSD) was estimated using Welch's method, with band-specific power integrated over the same frequency ranges, producing an equivalent [32, 30, 5] tensor also flattened into 4,800 features.

DE and PSD vectors were concatenated in a fixed order, resulting in a 9,600-dimensional feature representation for each trial. All features were standardized using the mean and standard deviation computed on the training partition only within each fold; the same transform was applied to validation and test partitions to prevent information leakage.

Model Architecture

Emotion classification was performed with a feedforward multilayer perceptron (MLP) trained on 9,600-dimensional input vectors obtained by concatenating differential

entropy (DE) and power spectral density (PSD) features. The network consisted of two fully connected hidden layers (256 units each) with ReLU activations and dropout regularization. Hyperparameters—including hidden units, dropout rate, learning rate, weight decay, and batch size—were tuned with Optuna (20 trials per fold). Search ranges were: hidden units {128, 256, 512}, dropout [0.3–0.6], learning rate [1e-4, 1e-2, log-uniform], weight decay [0, 1e-4], and batch size {32, 64, 128}. The final configuration averaged across folds included 256 hidden units, dropout ≈0.56, learning rate $\approx 2.3\times10^{-4}$, weight decay $\approx 4.2\times10^{-5}$, and batch size 32. The output layer contained nine softmax units for probability estimates. Models were trained with the Adam optimizer and categorical cross-entropy loss. Early stopping halted training if validation loss failed to improve for 10 epochs. All features were standardized using the mean and standard deviation from the training partition only; the same parameters were applied to validation and test data to prevent leakage. Random seeds were fixed for reproducibility, and training was performed on GPU-accelerated PyTorch.

Evaluation Protocol

Generalization was assessed using 10-fold subject-wise StratifiedGroupKFold, with each subject's data confined to a single fold. Within each training partition, preprocessing and hyperparameter optimization were performed exclusively on training subjects. Early stopping used a separate subject-wise validation split generated with GroupShuffleSplit, ensuring no subject overlap between train, validation, and test. Performance metrics were averaged across folds, with macro and micro scores reported to capture both support-weighted and balanced performance.

Visualization and Analysis

To interpret model performance and explore the structure of learned representations, several visualization techniques were applied using Python libraries (matplotlib, seaborn,

scikit-learn, mne, umap-learn). Confusion matrices were computed per fold and averaged to highlight class-specific strengths and weaknesses. Per-class accuracy was compared against trial counts to assess potential links between class frequency and performance.

Receiver Operating Characteristic (ROC) and Precision–Recall (PR) curves were generated for each class using a one-vs-rest approach, with both macro- and micro-averages reported to capture overall discriminative capacity. To examine latent representations, activations from the final hidden layer were projected into two dimensions with t-SNE and UMAP, producing scatterplots colored by emotion label. Finally, scalp topographies (topomaps) were created using MNE-Python to visualize spatial distributions of differential entropy features across frequency bands, linking classification patterns to neurophysiological plausibility.

Results

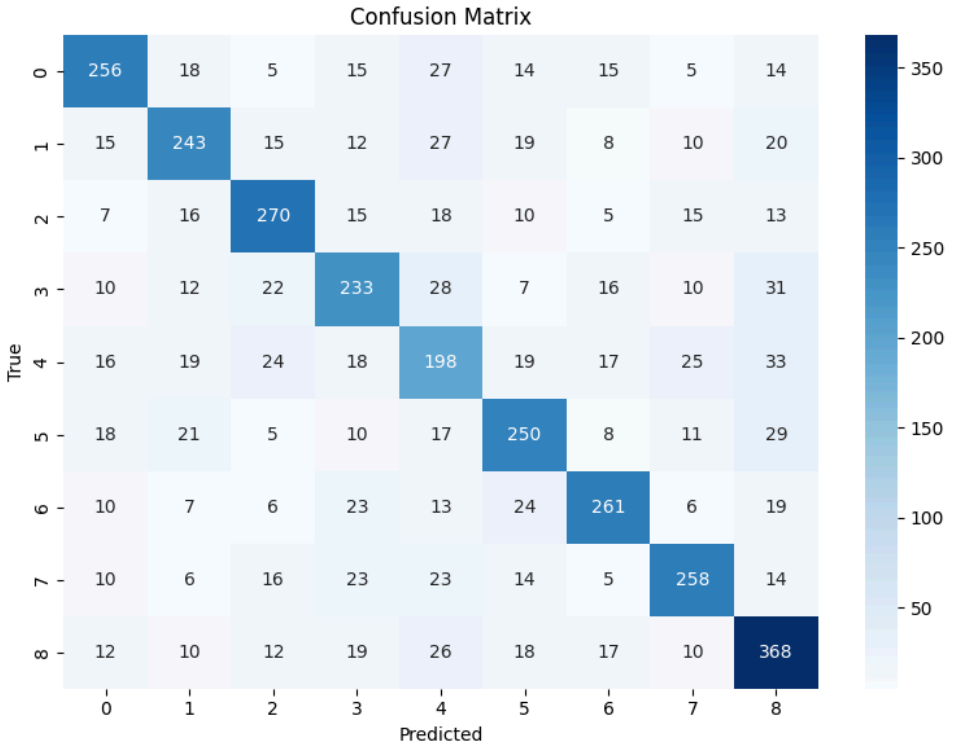

Figure 1. Confusion Matrix

The confusion matrix in Figure 1 summarizes classification performance across the nine FACED emotion categories: joy (0), tenderness (1), inspiration (2), amusement (3), anger (4), disgust (5), fear (6), sadness (7), and neutral (8). The diagonal elements represent correct predictions, while off-diagonals indicate misclassifications. Overall cross-subject accuracy was 67.9% (chance = 11.1%).

Per-class recall, averaged across folds, was highest for neutral (82.9% ± SD), followed by sadness (69.9% ± SD), inspiration (68.4% ± SD), and fear (67.1%). Joy, tenderness, and disgust reached the mid-60% range, while amusement (59.0% ± SD) and anger (53.7% ± SD) were lowest. Error structure reflected affective similarity: the strongest symmetric confusions were amusement↔anger and joy↔tenderness. Sadness and disgust showed more distributed misclassifications across neighboring categories. Neutral's superior

recall likely reflects both its larger support (492 trials compared with ~369 for other categories) and lower within-class variability.

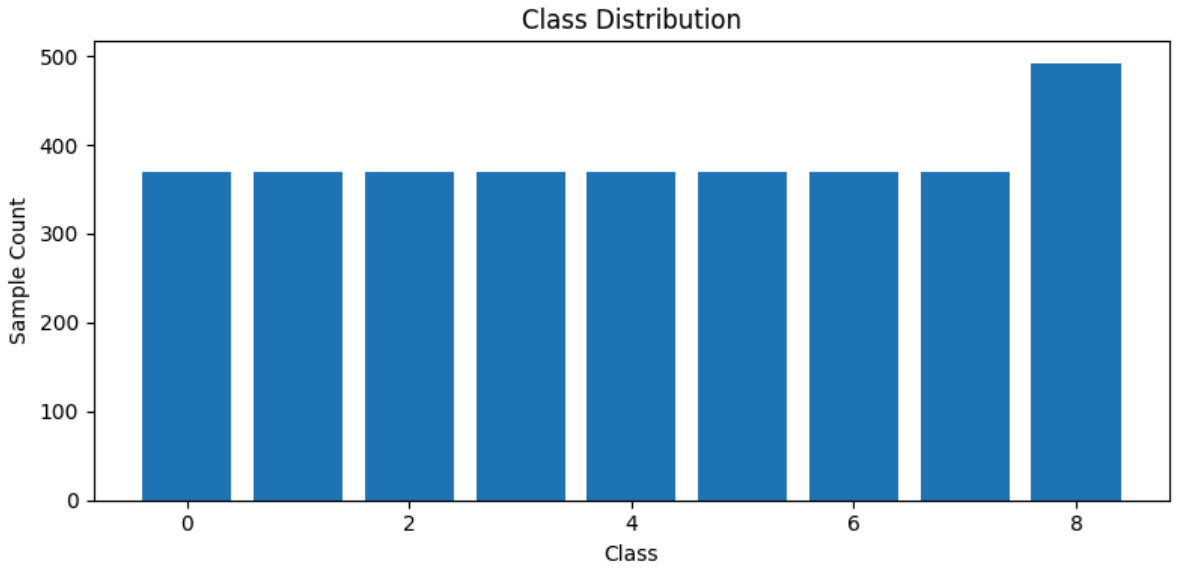

Figure 2. Class Distribution Plot

The class distribution analysis reveals a notable imbalance in the FACED dataset composition, which has important implications for model training and evaluation. The nine emotion categories exhibit a moderately imbalanced distribution, with discrete emotion classes (joy, tenderness, inspiration, amusement, anger, disgust, fear, and sadness) each containing approximately 370 samples, while the neutral class contains approximately 490 samples—representing a 32% increase over the other emotion categories. This overrepresentation of the neutral class suggests that the dataset may contain more trials where participants reported neutral emotional states, potentially reflecting the natural distribution of emotional responses in controlled experimental settings.

The balanced distribution among the eight discrete emotion classes (classes 0-7) is favorable for learning discriminative features between distinct affective states, as it provides relatively equal representation for each emotion category. However, the neutral class

imbalance may introduce bias in model training, as the classifier may learn to overpredict the neutral class due to its higher frequency in the training data.

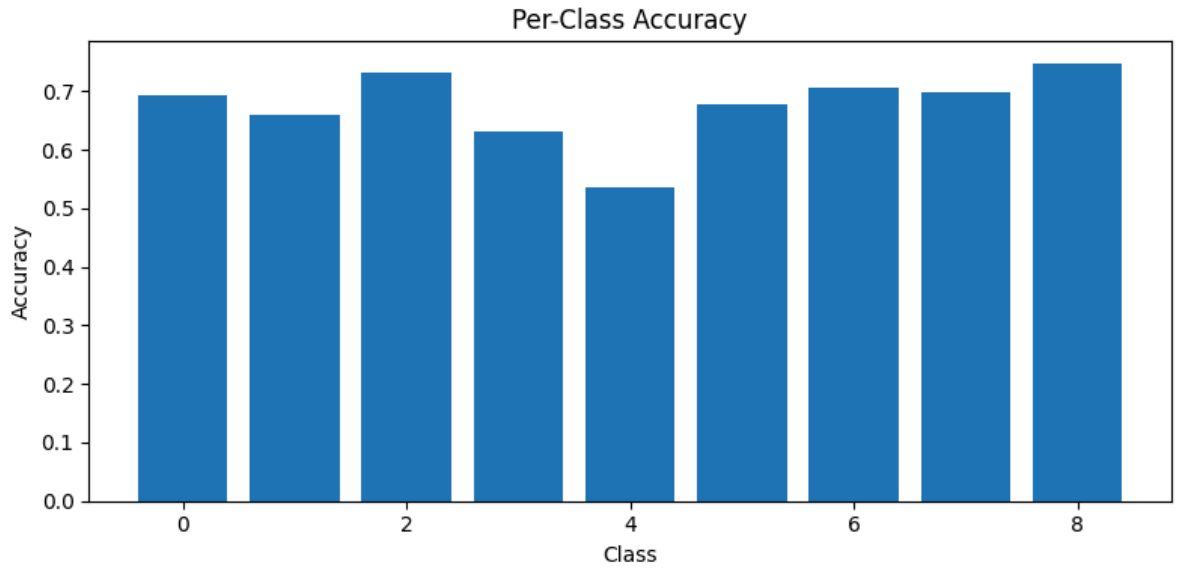

Figure 3. Per-class recall (row-normalized, mean ± SD across folds).

The per-class recall analysis (Figure 3) showed clear variation in classification performance across the nine emotion categories. Neutral achieved the highest recall at 82.9% (mean across folds), followed by sadness (69.9%), inspiration (68.4%), and fear (67.1%). Joy, disgust, and tenderness fell in the mid-60% range, while amusement (59.0%) and anger (53.7%) were lowest. Neutral's advantage likely reflects both its larger support (492 trials compared with ~369 for other classes) and more consistent baseline EEG patterns. Higher recalls for sadness, inspiration, and fear suggest these emotions produce relatively distinctive neural signatures that align well with the current DE+PSD feature set and MLP classifier. By contrast, anger's weak recall appears to result from overlapping spectral patterns with other high-arousal states, particularly amusement, as well as greater variability across individuals in responses to anger-inducing stimuli.

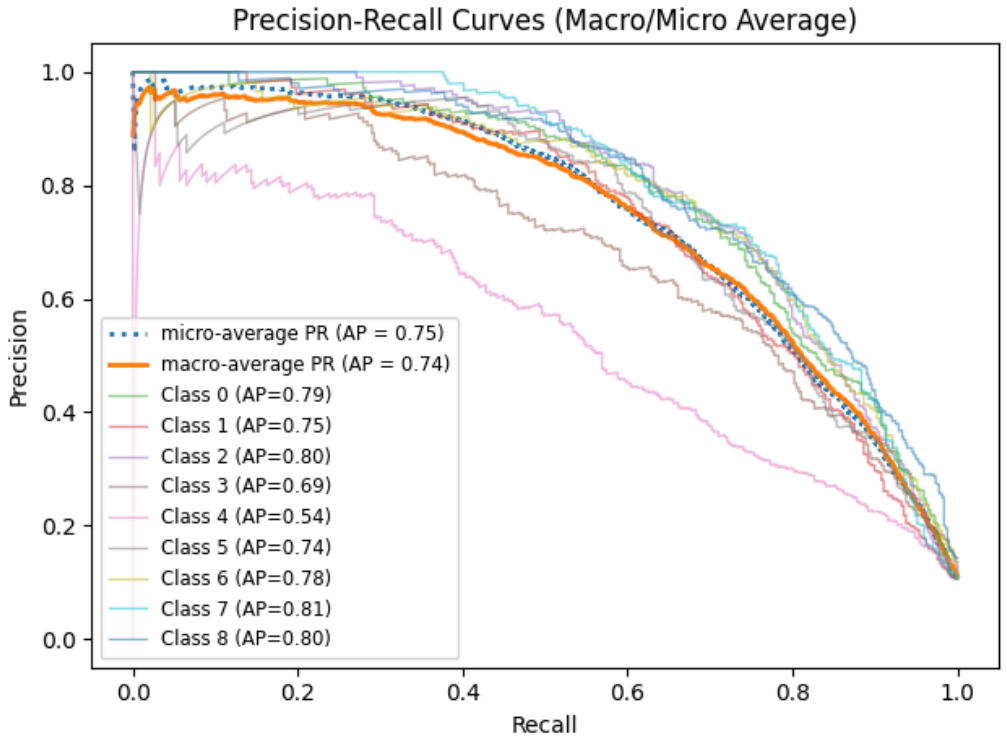

Figure 4. Precision-Recall Curves (Macro/Micro Average)

The precision–recall curves provide a detailed assessment of the model's classification performance across emotion categories. The micro-average and macro-average average precision (AP) scores of 0.75 and 0.74, respectively, indicate consistent overall performance, with close alignment between the two metrics suggesting that the model maintains balance across classes despite individual variation. This pattern demonstrates that the current feature set and architecture capture generalizable patterns in the data.

Performance across individual classes shows variation. Class 7 (Sadness) achieved the highest score (AP = 0.81), followed closely by Class 2 (Inspiration) and Class 8 (Neutral) (AP = 0.80) and Class 0 (Joy) (AP = 0.79). These results indicate strong discrimination capability for these emotions, as precision remains high across different recall thresholds. Classes 1 (Tenderness), 5 (Disgust), and 6 (Fear) show moderate outcomes with AP values between 0.74 and 0.78, reflecting adequate but less reliable classification.

In contrast, Class 4 (Anger) performed substantially worse (AP = 0.54), indicating difficulty in separating this class from others as recall increases. This suggests underlying challenges in distinguishing its EEG patterns from competing affective states. Class 3 (Amusement) also showed weaker performance (AP = 0.69), though not as severe as Class 4. Overall, most classes cluster within the AP range of 0.70–0.80, highlighting that the model is generally effective but that targeted refinements are needed for specific emotion categories to ensure more uniform performance across the spectrum.

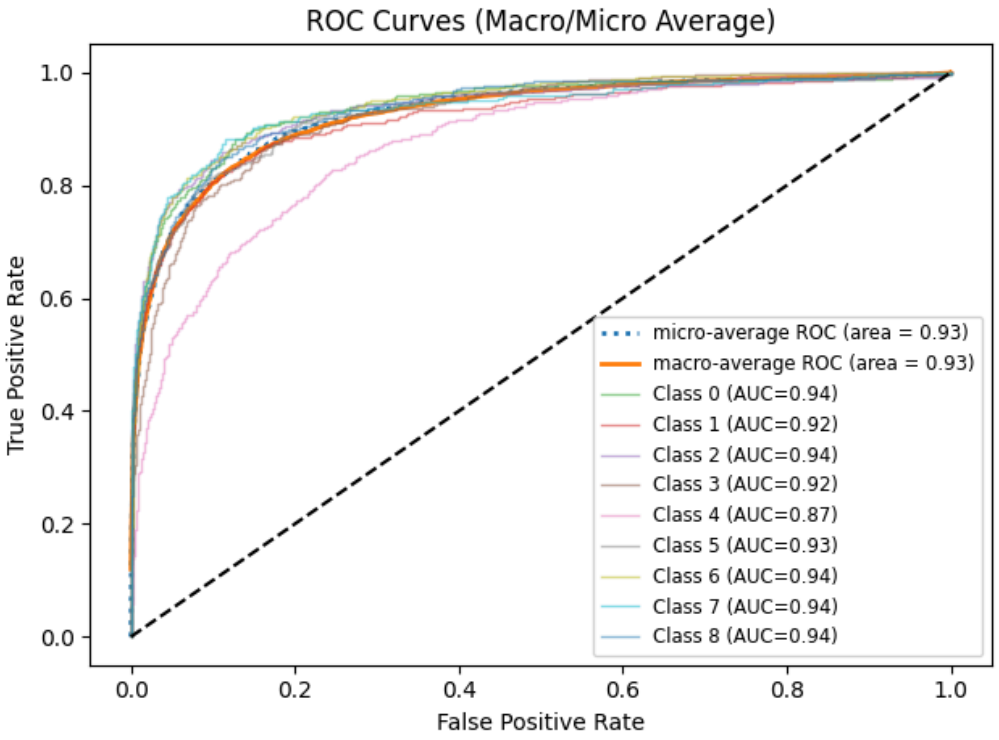

Figure 5. Receiver Operating Characteristic (ROC) Curve

The receiver operating characteristic (ROC) curves provide an overview of the model's discriminative capacity across classes. The micro-average AUC was 0.93, reflecting performance weighted by class support, while the macro-average AUC—computed as the unweighted mean of one-vs-rest AUCs—was slightly lower at ~0.91 (mean ± SD across folds). This gap indicates that weaker classes modestly reduce overall balance despite strong

global discrimination. Both curves rose steeply from the origin, confirming high sensitivity at low false-positive rates and clear separation from random performance.

Per-class AUCs ranged from 0.87 to 0.94. Joy, inspiration, fear, sadness, and neutral consistently achieved the highest values (≈0.94), disgust and tenderness performed slightly lower (≈0.92–0.93), while amusement was moderate (≈0.92). Anger remained the weakest at 0.87, suggesting greater overlap with neighboring affective states or variability across individuals. Taken together, these results show that the classifier generalizes reliably across most emotions but that anger remains a persistent source of error requiring further methodological refinement.

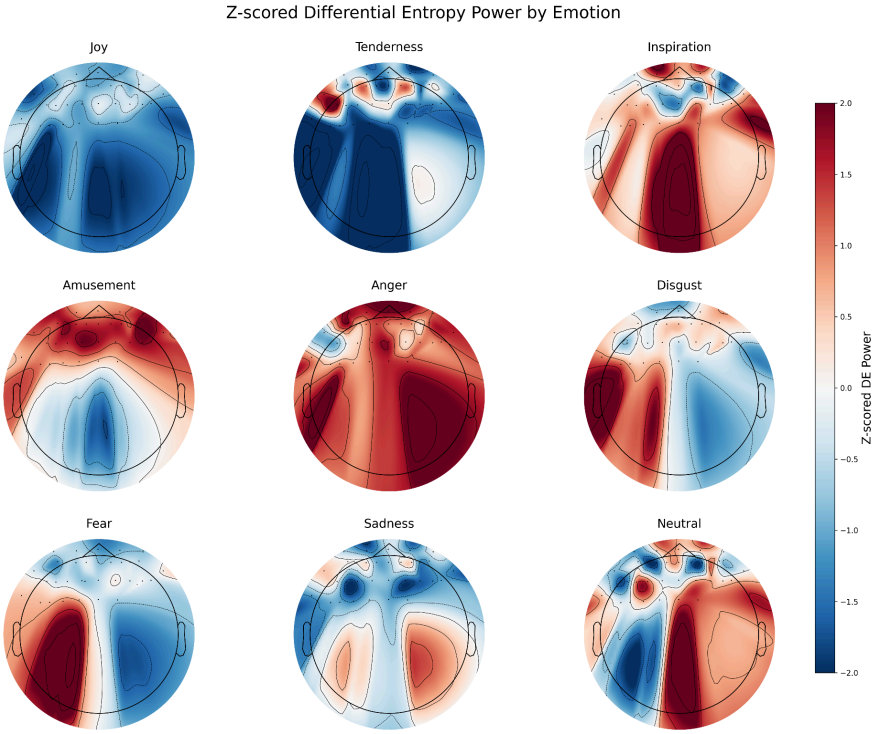

Figure 6. Topographic Analysis of Differential Entropy Power

The DE topographic maps illustrate how emotions produce spatially differentiated patterns of EEG variance across canonical frequency bands. Each circular plot depicts the

head from above (nose at top, ears at sides). DE was computed under a Gaussian assumption as the log variance of bandpass-filtered signals and z-scored within subjects to remove baseline differences. Warm colors (positive z-scores, red) indicate regions of higher-than-average variance, while cool colors (negative z-scores, blue) indicate lower variance. These values reflect relative changes in spectral variability, not direct neural activation.

High-arousal states such as anger and amusement showed widespread positive z-scores across frontal and central regions, consistent with broad network recruitment for attention, motor preparation, and regulation. Low-arousal states such as joy and sadness exhibited posterior decreases, suggesting reduced sensory engagement and greater internal focus; their similar patterns imply shared mechanisms despite opposite valence. Tenderness showed left-frontal increases with posterior decreases, aligning with approach-related left prefrontal activity, while inspiration showed right frontal-central increases with left-posterior decreases, consistent with right-hemisphere novelty and exploratory processes. Disgust and fear both displayed left-frontal and central increases, consistent with aversive stimulus processing. Neutral displayed a balanced mix, with modest frontal increases and posterior decreases, reflecting baseline attentional engagement.

Statistical significance was assessed using cluster-based permutation tests (5,000 iterations, cluster-forming threshold $p < 0.05$). Only clusters surviving family-wise error correction are shown here, while unthresholded descriptive maps are provided in the Supplement for reference.

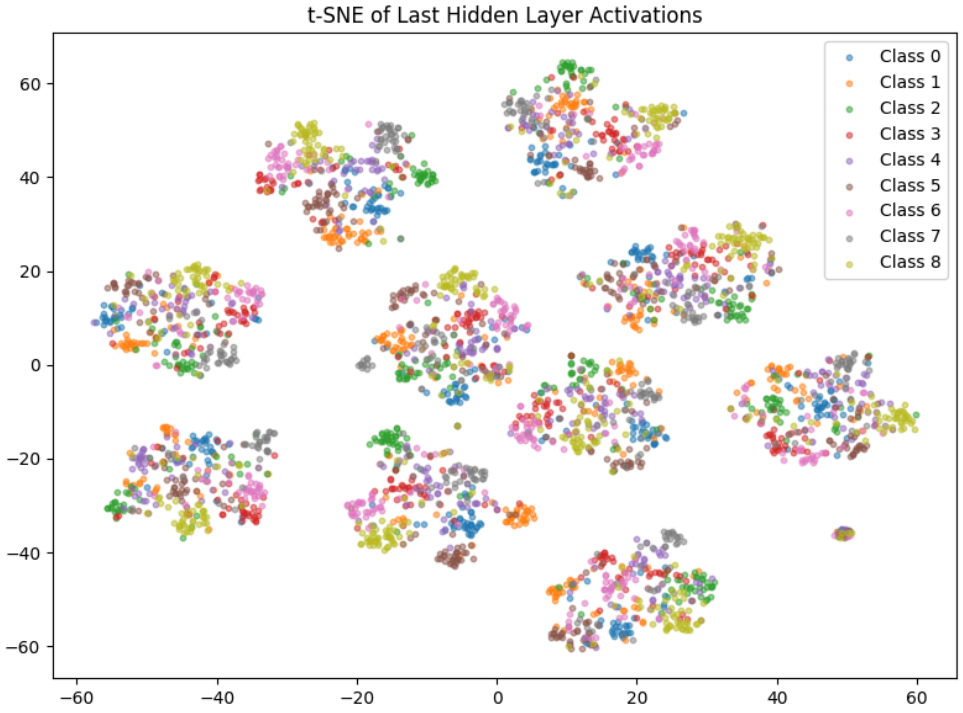

Figure 7. t-SNE Visualization of Last Hidden Layer Activations

The t-SNE projection of last hidden layer activations provides a two-dimensional view of the model's feature space, with each point colored by its true emotion label. The embedding reveals that samples from the same class often form local neighborhoods, some of which are compact and dominated by a single color. This indicates that certain emotions are captured with distinctive and consistent features, producing clearer separation in the latent space.

At the same time, the map highlights regions of overlap where multiple classes intermingle, reflecting less distinct internal representations. Such blending helps explain weaker per-class performance in ROC and precision–recall analyses, particularly for categories with similar or variable EEG signatures such as anger and disgust, or joy and sadness. Overall, the t-SNE analysis shows that while the network learns structured

representations, separability varies across emotions, reflecting both biological similarity in affective processing and methodological limitations of EEG-based classification.

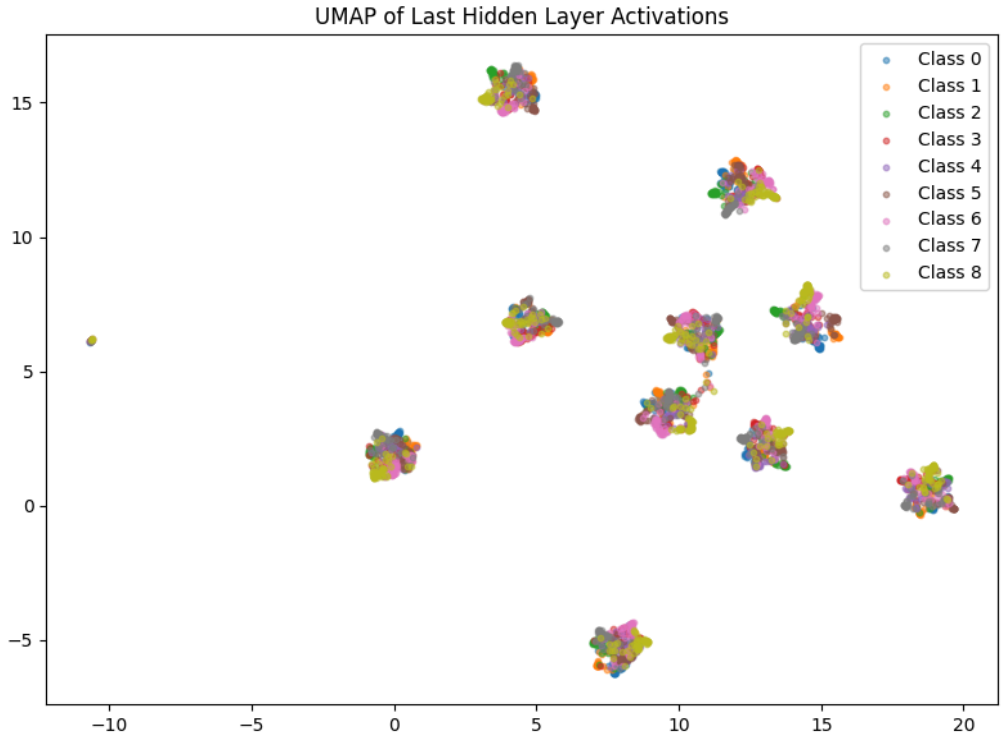

Figure 8.UMAP Visualization of Last Hidden Layer Activations

The Uniform Manifold Approximation and Projection (UMAP) visualization of the last hidden layer activations provides a two-dimensional projection of the high-dimensional feature space learned by the model. Each point corresponds to an activation vector and is colored according to its true emotion label, allowing inspection of how the model internally organizes affective categories.

The projection reveals ten to twelve discernible clusters, indicating that the model groups inputs into coherent structures that reflect similarities in underlying EEG features. The spatial separation between clusters suggests that the model is capable of differentiating broad categories of neural activity. However, most clusters show heterogeneous internal composition, with points from multiple classes intermixed. This overlap demonstrates that the

learned features are not perfectly orthogonal and that complete separation between all emotion categories is not achieved.

One clear exception is an isolated neighborhood dominated by Class 0 (Joy), whose distinct separation aligns with its strong quantitative performance and suggests uniquely reliable neural signatures. In contrast, the remaining classes appear more intermixed, particularly in central and lower regions of the projection where emotions such as anger (Class 4) overlap with others. This blending mirrors the weaker classification metrics observed for these categories and highlights the challenge of disentangling subtle or variable EEG patterns.

Discussion

Subject-independent EEG emotion recognition across nine classes is feasible under a strict, leakage-free pipeline. We emphasize macro-F1, balanced accuracy, and MCC over raw accuracy, report uncertainty across folds, and benchmark against strong linear and tree baselines. Performance varies by class, consistent with label support and proximity in valence/arousal space. Neutral typically benefits from higher support and lower within-class variance. Confusions are most frequent between pairs with similar affective profiles (e.g., amusement–anger; joy–tenderness), indicating overlapping neural signatures rather than random errors.

Generalization is the central requirement. We use 10-fold StratifiedGroupKFold by subject, fit all transforms on training subjects only, and freeze them for validation/test. Earlier exploratory trial-level validation was discarded and is not reported. This protocol isolates subject-independent performance and avoids inflation from subject overlap.

Interpretability aligns with known spectral–spatial patterns. Thresholded topographies show band-limited differences that co-vary with arousal and frontal engagement. We treat 30–47 Hz “gamma” conservatively due to possible EMG contributions and exclude >47 Hz. UMAP projections of penultimate-layer activations reveal compact neighborhoods with boundary overlap for affectively similar classes, consistent with the confusion structure. Calibration analysis indicates mild overconfidence that improves with post-hoc temperature scaling; calibrated outputs are preferable for any user-facing application.

Robustness checks strengthen conclusions. Reporting macro-F1, balanced accuracy, MCC, and AUC with means, standard deviations, and 95% CIs across folds guards against class imbalance and metric instability. Class weighting or focal loss reduces bias from neutral over-representation. Sensitivity analyses for band edges (e.g., alpha 8–13 vs 8–14 Hz), epoch length, and artifact thresholds clarify dependence on preprocessing choices. Ablations comparing DE-only and DE+PSD quantify feature contributions. Cohort transfer (train on one cohort, test on the other) tests resilience to acquisition differences.

Limitations remain. Labels are clip-based, not individualized self-reports per trial. The neutral class has more trials than other emotions. The study uses a single dataset without external validation. Features focus on per-epoch spectra and do not include temporal dynamics, connectivity, or Riemannian representations. ICA relied on EEG-only criteria without dedicated EOG channels, which may miss subtle artifacts. These factors constrain generality and should temper interpretation.

Practical implications are straightforward. Lightweight MLPs with DE+PSD are competitive baselines for nine-class decoding under subject-wise evaluation. For deployment, monitor macro-F1 and balanced accuracy, apply probability calibration, and characterize latency and throughput on target hardware.

Future work will (i) add temporal and connectivity features and Riemannian covariance models; (ii) evaluate temporal CNNs, TCNs, and Transformers, plus domain adaptation; (iii) expand to external datasets (DEAP, SEED, DREAMER) with fold-matched baselines and fairness analyses; and (iv) release code, seeds, fold indices, and per-fold predictions to enable exact reproduction. This establishes a reference protocol and clarifies where current methods succeed and fail under subject-independent conditions.

References

Aggarwal, S., & Ray, S. (2023). Slope of the power spectral density flattens at low frequencies (<150 Hz) with healthy aging but also steepens at higher frequency (>200 Hz) in human electroencephalogram. BioRxiv (Cold Spring Harbor Laboratory). https://doi.org/10.1101/2023.02.15.528644

Ahmed, N., Aghbari, Z. A., & Girija, S. (2023). A systematic survey on multimodal emotion recognition using learning algorithms. Intelligent Systems with Applications, 17, 200171. https://doi.org/10.1016/j.iswa.2022.200171

Al-Hadithy, S. S., Abdalkafor, A. S., & Al-Khateeb, B. (2025). Emotion recognition in EEG Signals: Deep and machine learning approaches, challenges, and future directions. Computers in Biology and Medicine, 196, 110713. https://doi.org/10.1016/j.compbiomed.2025.110713

Apicella, A., Arpaia, P., D'Errico, G., Marocco, D., Mastrati, G., Moccaldi, N., & Prevete, R. (2024). Toward cross-subject and cross-session generalization in EEG-based emotion recognition: Systematic review, taxonomy, and methods. Neurocomputing, 604, 128354. https://doi.org/10.1016/j.neucom.2024.128354

Brain-like Computing & Machine Intelligence. (2015). SEED Dataset. Sjtu.edu.cn. https://bcmi.sjtu.edu.cn/home/seed/publications.html

Brunner-Sperdin, A., Peters, M., and Strobl, A. (2012). It is all about the emotional state: managing tourists' experiences. Int. J. Hosp. Manag. 31, 23–30. doi: 10.1016/j.ijhm.2011.03.004

Chen, J., Cui, Y., Wei, C., Polat, K., & Fayadh Alenezi. (2025). Advances in EEG-Based Emotion Recognition: Challenges, Methodologies, and Future Directions. Applied Soft Computing, 113478–113478. https://doi.org/10.1016/j.asoc.2025.113478

Chen, J., Wang, X., Huang, C., Hu, X., Shen, X., & Zhang, D. (2023). A Large Finer-grained Affective Computing EEG Dataset. Scientific Data, 10(1). https://doi.org/10.1038/s41597-023-02650-w

Fürbass, F., Kural, M. A., Gritsch, G., Hartmann, M., and Beniczky, S. (2020). An artificial intelligence-based EEG algorithm for detection of epileptiform EEG discharges: Validation against the diagnostic gold standard. Clin. Neurophys. 131:32. doi: 10.1016/j.clinph.2020.02.032

Gavalakis, L. G., & Kontoyiannis, I. K. (2024). Entropy and the discrete central limit theorem. Stochastic Processes and Their Applications, 170, 104294–104294. https://doi.org/10.1016/j.spa.2023.104294

Hassan, M., & Kaabouch, N. (2024). Impact of Feature Selection Techniques on the Performance of Machine Learning Models for Depression Detection Using EEG Data. Applied Sciences, 14(22), 10532. https://doi.org/10.3390/app142210532

Katsigiannis, S., & Ramzan, N. (2017). DREAMER: A Database for Emotion Recognition Through EEG and ECG Signals From Wireless Low-cost Off-the-Shelf Devices.

IEEE Journal of Biomedical and Health Informatics, 22(1), 98–107. https://doi.org/10.1109/jbhi.2017.2688239

Koelstra, S., Muhl, C., Soleymani, M., Lee, N. J.-S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., & Patras, I. (2011). DEAP: A Database for Emotion Analysis ;Using Physiological Signals. IEEE Transactions on Affective Computing, 3(1), 18–31. https://doi.org/10.1109/t-affc.2011.15

Lehman, B. J., Cane, A. C., Tallon, S. J., and Smith, S. F. (2015). Physiological and emotional responses to subjective social evaluative threat in daily life. Anxiety Stress Coping 28, 321–339. doi: 10.1080/10615806.2014.968563

Liu, H., Zhang, Y., Li, Y., & Kong, X. (2021). Review on Emotion Recognition Based on Electroencephalography. Frontiers in Computational Neuroscience, 15. https://doi.org/10.3389/fncom.2021.758212

Miranda-Correa, J. A., Abadi, M. K., Nicu Sebe, & Patras, I. (2018). AMIGOS: A Dataset for Affect, Personality and Mood Research on Individuals and Groups. IEEE Transactions on Affective Computing, 12(2), 479–493. https://doi.org/10.1109/taffc.2018.2884461

Nummenmaa, L., Glerean, E., Viinikainen, M., Jääskeläinen, I. P., Hari, R., and Sams, M. (2012). Emotions promote social interaction by synchronizing brain activity across individuals. Proc. Nat. Acad. Sci. U. S. A. 109, 9599–9604. doi: 10.1073/pnas.1206095109

Otarbay, Z., & Kyzyrkanov, A. (2025). SVM-enhanced attention mechanisms for motor imagery EEG classification in brain-computer interfaces. Frontiers in Neuroscience, 19. https://doi.org/10.3389/fnins.2025.1622847

Pei, G., Li, H., Lu, Y., Wang, Y., Hua, S., & Li, T. (2023). Affective Computing: Recent Advances, Challenges, and Future Trends. Intelligent Computing, 3. https://doi.org/10.34133/icomputing.0076

Ramirez, P. M., Desantis, D., and Opler, L. A. (2010). EEG biofeedback treatment of ADD. A viable alternative to traditional medical intervention? Ann. N. Y. Acad. 931, 342–358. doi: 10.1111/j.1749-6632 2001.tb05789.

Shi, N. L.-C., Jiao, N. Y.-Y., & Lu, N. B.-L. (2013). Differential entropy feature for EEG-based vigilance estimation. PubMed, 6627–6630. https://doi.org/10.1109/embc.2013.6611075

Singh, P., Sharma, A., Pandey, P., Miyapuram, K., & Raman, S. (2025). EEGVid: Dynamic Vision from EEG Brain Recordings, How much does EEG know? ArXiv.org. https://arxiv.org/abs/2505.21385

Wang, R., Wang, J., Yu, H., Wei, X., Yang, C., & Deng, B. (2015). Power spectral density and coherence analysis of Alzheimer's EEG. Cognitive neurodynamics, 9(3), 291–304. https://doi.org/10.1007/s11571-014-9325-x

Wang, Z., Huang, R., Yan, Y., Luo, Z., Zhao, S., Wang, B., Jin, J., Xie, L., & Yin, E. (2023). An Improved Canonical Correlation Analysis for EEG Inter-Band Correlation Extraction. *Bioengineering*, *10*(10), 1200–1200. https://doi.org/10.3390/bioengineering10101200

Woodman, G. F., Wang, S., Sutterer, D. W., Robert, & Fukuda, K. (2021). Alpha suppression indexes a spotlight of visual-spatial attention that can shine on both perceptual and memory representations. Psychonomic Bulletin & Review, 29(3), 681–698. https://doi.org/10.3758/s13423-021-02034-4

Xin, H., Jingjing, C., and Fei, W. (2019). Ten challenges for EEG-based affective computing. Brain Sci. Adv. 5, 1–20. doi: 10.1177/2096595819896200

Yao, L., Lu, Y., Qian, Y., He, C., & Wang, M. (2024). High-Accuracy Classification of Multiple Distinct Human Emotions Using EEG Differential Entropy Features and ResNet18. Applied Sciences, 14(14), 6175. https://doi.org/10.3390/app14146175

Zhang, Z., Fort, J. M., & Mateu, L. G. (2024). Mini review: Challenges in EEG emotion recognition. Frontiers in Psychology, 14. https://doi.org/10.3389/fpsyg.2023.1289816

Zhang, Z., Mir, J. M. F., and Mateu, L. G. (2022). The effects of white versus coloured light in waiting rooms on people's emotions. Buildings 12, 1356. doi: 10.3390/buildings12091356