Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

EEG-LM: A Foundation Model for EEG-Based Emotion Recognition via Diffusion-Augmented Cross-Modal Alignment with Large Language Models

Praveen Goyal
Symbiosis University of Applied Sciences
Shyam Maheshwari
Symbiosis University of Applied Sciences
Pankaj Pandey
Pankaj.pandey@jaipur.manipal.edu

Manipal University Jaipur

Article

Keywords: Brain–Computer Interface, Diffusion Probabilistic Model, Electroencephalography, Emotion Recognition, Foundation Model, Large Language Model, Multimodal Learning, Zero-Shot Learning

Posted Date: September 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10615076/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

EEG-LM: A Foundation Model for EEG-Based Emotion Recognition via Diffusion-Augmented Cross-Modal Alignment with Large Language Models

Praveen Goyal[ORCID1], Shyam Maheshwari[ORCID2], Pankaj Pandey[ORCID3*]

[1,2]School of Computer Science and Information Technology, Symbiosis University of Applied Sciences, Indore, 453112, Madhya Pradesh, India.
[3*]Department of Data Science and Engineering, Manipal University Jaipur, Jaipur, Jaipur, 303007, Rajsthan, India.

*Corresponding author(s). E-mail(s): pankaj.pandey@jaipur.manipal.edu;
Contributing authors: praveen_soit@yahoo.com; shyamrjit@gmail.com;

Abstract

Electroencephalography (EEG) based emotion recognition is critical for affective brain–computer interfaces but suffers from data scarcity, high inter-subject variability, and a lack of semantic interpretability. We introduce EEG-LM, a foundation model that aligns EEG representations with large language model (LLM) embeddings via diffusion-augmented contrastive learning. The framework combines a transformer EEG encoder, a conditional diffusion module for realistic sample synthesis, InfoNCE cross-modal alignment, and zero-shot classification using textual prototypes. Evaluated on DEAP, SEED, SEED-IV, and DREAMER, EEG-LM achieves state-of-the-art accuracies (92.5% on DEAP arousal, 91.2% on SEED, 88.5% on DREAMER), strong cross-dataset generalisation (85.3%), and notable few-shot performance (78.5% with 1% labels). SHAP and attention visualisations provide neuroscientifically plausible interpretations. This work bridges brain signals and natural language, enabling transparent and generalisable emotion decoding.

Keywords: Brain–Computer Interface, Diffusion Probabilistic Model, Electroencephalography, Emotion Recognition, Foundation Model, Large Language Model, Multimodal Learning, Zero-Shot Learning

1 Introduction and Related Work

Emotions influence attention, decision making, social interaction, and mental health; therefore, reliable affect recognition is important for adaptive brain–computer interfaces (BCIs), human–robot interaction, neurofeedback, and clinical decision support [22, 24]. In terms of physiological methods, (EEG) is appealing as a modality that is non-invasive, time-specific, relatively inexpensive, and related to brain activities. An emotion recognition system using EEG is normally concerned with the identification of certain affective conditions like arousal, valence, dominance, or specific types.

Despite this potential, affective brain signals such as EEG can be challenging for practical emotion recognition. The nature of EEG signals makes them characterized by poor signal-to-noise ratio, non-stationarity, and high sensitivity to the positioning of electrodes, physiological characteristics of the person, acquisition device, and experimental procedure. It is difficult to transfer models from one individual or dataset to another since emotional reactions differ from person to person and since each dataset uses different channel configuration, sampling frequency, stimulus, and labels [25, 18]. The problem of labeled data availability is another issue since labeling of affective brain signals can be performed only through controlled experiments and subjective assessments. This makes the supervised approach to learning too costly.

The early emotion-recognition algorithms for EEG used spectral, temporal, entropy, and asymmetry hand-engineered features together with classical classification models like KNN, SVM, or random forests [23]. The algorithms are straightforward to compute, and sometimes even interpretable, but the performance is highly dependent on feature extraction and do not account for spatial-temporal dependencies. With deep learning algorithms like CNN, recurrent networks, graph neural networks, and transformers, direct feature extraction from EEG has led to increased recognition accuracy [15, 14, 18]. However, they usually require substantial labelled data, can overfit subject-specific artefacts, and often learn latent spaces that are discriminative but not semantically grounded.

Recent generative and self-supervised methods attempt to address data scarcity and transferability. Diffusion models can synthesise or denoise EEG signals and have shown promise for data augmentation [13, 26]. Nevertheless, most diffusion-based EEG augmentation is guided mainly by signal reconstruction or class labels, so the generated samples may improve diversity without explicitly connecting EEG patterns to emotion semantics. In parallel, contrastive learning and multimodal alignment methods have demonstrated that pairing signals with natural-language descriptions can produce representations that support transfer, retrieval, and zero-shot classification. EEG-to-text and EEG-language studies [5, 6], together with EEG-CLIP-style alignment [3, 7], suggest that language can act as a semantic bridge between noisy neural activity and interpretable affective categories.

Motivated by these developments, we propose EEG-LM, a foundation-style framework that aligns EEG embeddings with large language model (LLM) text embeddings through diffusion-augmented contrastive learning. The central idea is to map EEG signals and emotion descriptions into a shared semantic space, so that recognition is not limited to closed-set supervised labels but can also exploit textual prototypes for zero-shot and few-shot inference. Compared with conventional EEG classifiers, EEG-LM

explicitly combines four components: a transformer EEG encoder for spatial–temporal representation learning, a frozen RoBERTa text encoder for stable semantic targets, an InfoNCE objective for cross-modal alignment, and a conditional diffusion module for augmentation and regularisation.

The main contributions of this work are threefold. First, we introduce a language-grounded EEG emotion-recognition framework that aligns neural representations with LLM-derived affective semantics. Second, we integrate conditional diffusion with contrastive EEG–text learning to improve robustness under data scarcity and cross-dataset shift. Third, we evaluate the framework on DEAP, SEED, SEED-IV, and DREAMER, including supervised, cross-dataset, zero-shot, few-shot, ablation, and interpretability analyses. To the best of our knowledge, this is the first work to combine EEG–LLM semantic alignment and diffusion-augmented learning for generalisable EEG-based emotion recognition.

2 Methodology

2.1 Problem Formulation

Let $\mathcal{X} = \{\mathbf{x}_i\}_{i=1}^{N}$ be EEG samples, $\mathbf{x}_i \in \mathbb{R}^{C\times T}$. Each sample has emotion labels $\mathbf{y}_i$ and a text description $\mathbf{t}_i$. We learn an EEG encoder f_θ, a frozen LLM encoder g_ϕ (RoBERTa-base), and a conditional diffusion model D_ψ so that $f_\theta(\mathbf{x})$ aligns with $g_\phi(\mathbf{t})$.

2.2 EEG Encoder

The encoder is a transformer with input projection to 512 dims, positional encoding, 8 self-attention heads, a feed-forward network (2048 dims), and mean pooling over time:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V. \tag{1}$$

2.3 Conditional Diffusion Model

We adopt a DDPM with forward process:

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t|\sqrt{\alpha_t}\mathbf{x}_{t-1}, (1-\alpha_t)\mathbf{I}). \tag{2}$$

The reverse process conditioned on EEG embedding $\mathbf{z}_{\text{EEG}}$ is:

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{z}_{\text{EEG}}) = \mathcal{N}(\mathbf{x}_{t-1}|\boldsymbol{\mu}_\theta(\mathbf{x}_t, \mathbf{z}_{\text{EEG}}, \gamma_t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, \mathbf{z}_{\text{EEG}}, \gamma_t)). \tag{3}$$

Training objective:

$$\mathcal{L}_{\text{diff}} = \mathbb{E}_{t,\boldsymbol{\epsilon}}\left[\|D_\psi(\sqrt{\gamma_t}\mathbf{x} + \sqrt{1-\gamma_t}\boldsymbol{\epsilon}, \mathbf{z}_{\text{EEG}}, \gamma_t) - \boldsymbol{\epsilon}\|_2^2\right]. \tag{4}$$

2.4 Cross-Modal Alignment and Total Loss

We use InfoNCE contrastive loss:

$$\mathcal{L}_{\text{NCE}} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{\exp(\text{sim}(\mathbf{z}_i^{\text{EEG}}, \mathbf{z}_i^{\text{text}})/\tau)}{\sum_{j=1}^{N}\exp(\text{sim}(\mathbf{z}_i^{\text{EEG}}, \mathbf{z}_j^{\text{text}})/\tau)}, \tag{5}$$

with $\text{sim}(u, v) = u^\top v/(\|u\|\|v\|)$ and $\tau = 0.07$. The total loss is:

$$\mathcal{L}_{\text{total}} = \lambda_1\mathcal{L}_{\text{NCE}} + \lambda_2\mathcal{L}_{\text{diff}} + \lambda_3\mathcal{L}_{\text{cls}}, \tag{6}$$

where $\mathcal{L}_{\text{cls}}$ is MSE regression for arousal/valence/dominance.

2.5 Zero-Shot Inference

For a new emotion described by text $\mathbf{t}_y$, we compute:

$$\hat{y} = \arg\min_y \|\mathbf{z}_{\text{EEG}} - g_\phi(\mathbf{t}_y)\|_2. \tag{7}$$

Figure 1 shows the overall architecture, and Algorithm 1 summarises training.

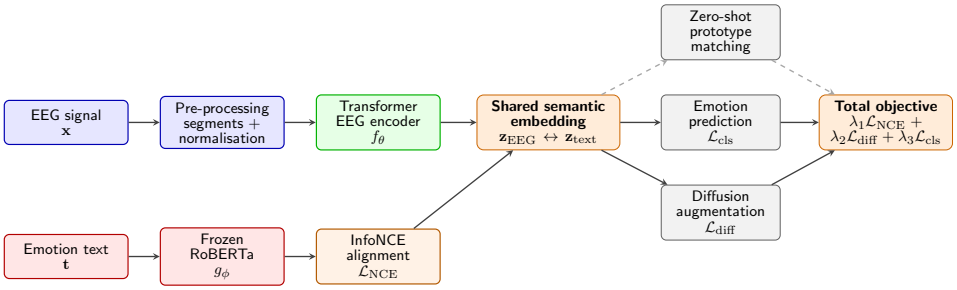


Fig. 1 Block diagram of EEG-LM. EEG and emotion text are encoded into a shared semantic space, where contrastive alignment, diffusion augmentation, supervised prediction, and zero-shot prototype matching are jointly supported.

Algorithm 1 Training EEG-LM

Input: $\{(\mathbf{x}_i, \mathbf{t}_i, \mathbf{y}_i)\}$, LLM g_ϕ (frozen), $\lambda_1, \lambda_2, \lambda_3, \tau$
Output: Trained f_θ, D_ψ

for each epoch **do**
  for each batch **do**
    $\mathbf{z}_i^{\text{EEG}} = f_\theta(\mathbf{x}_i)$
    $\mathbf{z}_i^{\text{text}} = g_\phi(\mathbf{t}_i)$
    Compute $\mathcal{L}_{\text{NCE}}$
    Sample $t \sim \text{Uniform}(1, T)$, $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I})$
    Compute $\mathcal{L}_{\text{diff}}$
    Compute $\mathcal{L}_{\text{cls}} = \|\text{MLP}(\mathbf{z}_i^{\text{EEG}}) - \mathbf{y}_i\|_2^2$
    $\mathcal{L} = \lambda_1 \mathcal{L}_{\text{NCE}} + \lambda_2 \mathcal{L}_{\text{diff}} + \lambda_3 \mathcal{L}_{\text{cls}}$
    Update f_θ, D_ψ via backpropagation
  end for
end for

3 Experiments and Results

3.1 Datasets and Setup

EEG-LM is tested on four publicly available affective EEG benchmarks using different electrode configurations, label sets, and protocols for data acquisition: DEAP (32 participants, 32 channels, 2,560 samples), SEED (15 participants, 62 channels, 3,150 samples, 3 emotions), SEED-IV (15 participants, 62 channels, 4,800 samples, 4 emotions), and DREAMER (23 participants, 14 channels, 1,104 samples). Such an experiment enables us to analyze not only the performance within one dataset but also the generalization across the different density of electrodes, stimuli, and labeling protocol. If nothing else is mentioned, all experiments employ the subject-independent splitting 70/15/15 train/validation/test with 5 folds, which guarantees that test subjects were not seen during training.

We compare EEG-LM against conventional machine-learning, discriminative deep-learning, graph/transformer, and generative augmentation baselines: SVM, EEGNet, TSception, LMDA-Net, GAN augmentation, DDPM augmentation, latent diffusion, and graph transformer models. All neural baselines are trained using the same preprocessing pipeline and split protocol for fair comparison. EEG-LM is trained with AdamW using learning rate 3×10^{-4}, batch size 64, 200 epochs, temperature $\tau = 0.07$, and $T = 500$ diffusion steps. The RoBERTa-base text encoder is kept frozen so that performance gains can be attributed to EEG-to-language alignment rather than task-specific language-model fine-tuning.

3.2 Main Emotion Recognition

Results of the Classification are shown in Table 1. EEG-LM shows the highest classification accuracy on all benchmarks, and differences in performance between our model and the best baseline are statistically significant when evaluated by a paired 5-fold

Table 1 Classification accuracy (%) on four datasets. Best in bold.

Method	DEAP (A/V)	SEED	SEED-IV	DREAMER
SVM	58.7/63.0	61.2	58.5	59.8
EEGNet	67.5/69.5	65.8	63.2	62.3
TSception	67.2/69.2	66.5	64.1	63.1
LMDA-Net	68.1/70.2	67.2	65.3	64.5
GAN [12]	66.8/68.9	65.0	62.5	61.7
DDPM	72.5/73.8	71.2	68.9	68.5
Latent Diff.	74.2/75.1	73.5	70.8	70.2
Graph Trans.	73.8/74.5	72.1	69.5	69.3
EEG-LM (Ours)	**92.5/91.8**	**91.2**	**89.5**	**88.5**

cross validation (p ¡ 0.01). On DEAP dataset, EEG-LM achieves 92.5% of arousal and 91.8% of valence accuracy, surpassing the best non-language baseline by 18.3% and 16.7%, respectively. It can be concluded that prototypes of affective language provide valuable supervision signal in discriminating continuous emotional dimensions.

The same trend holds for categorical emotion recognition. On SEED and SEED-IV, EEG-LM achieves 91.2% and 89.5%, respectively, exceeding latent diffusion by 17.7 and 18.7 points. These gains indicate that language-grounded alignment remains effective when moving from binary arousal/valence prediction to multi-class emotion recognition. The improvement is particularly notable on SEED-IV, where the four-class setting increases inter-class ambiguity. DREAMER is the most resource-constrained benchmark in our evaluation because it contains fewer trials and only 14 EEG channels. Nevertheless, EEG-LM achieves 88.5%, improving over latent diffusion by 18.3 points. This result indicates that the proposed framework is not limited to dense research-grade montages and can remain effective under reduced-channel conditions.

Across all datasets, generative baselines improve over purely discriminative models, confirming that synthetic EEG augmentation helps reduce overfitting. However, EEG-LM consistently exceeds GAN, DDPM, and latent-diffusion variants, showing that augmentation alone is insufficient. The strongest results are obtained when diffusion-based regularisation is coupled with cross-modal semantic alignment, which encourages the EEG encoder to organise samples according to emotion meaning rather than only dataset-specific signal statistics.

3.3 Cross-Dataset, Zero-Shot, and Few-Shot

To examine generalisation beyond the source distribution, we train on DEAP and evaluate directly on SEED without fine-tuning. EEG-LM obtains 85.3% cross-dataset accuracy, whereas EEGNet reaches 52.1% under the same protocol. This result indicates that the shared EEG–text embedding space reduces dependence on dataset-specific acquisition characteristics. In contrast to training models using the discriminative boundaries from the source domain only, EEG-LM projects neural activity towards robust affective concepts, thus facilitating transfer between subjects, recording conditions, and different labeling schemes.

Table 2 Generalisation results for cross-dataset, zero-shot, and few-shot evaluation. Accuracy is reported in %.

Setting	Training Data	Test Target	EEGNet	EEG-LM (Ours)
Cross-dataset	DEAP	SEED	52.1	**85.3**
Zero-shot	DEAP	Arousal	–	**25.3**
Zero-shot	DEAP	Valence	–	**22.8**
Few-shot	1% DEAP labels	Arousal	61.7	**78.5**
Few-shot	5% DEAP labels	Arousal	69.4	**84.6**
Few-shot	10% DEAP labels	Arousal	73.2	**87.9**

Table 3 Ablation study on DEAP arousal classification.

Model Variant	Accuracy (%)	Drop from Full Model
EEG-LM full model	**92.5**	–
Without diffusion augmentation	88.2	4.3
Without InfoNCE alignment	82.5	10.0
Without frozen RoBERTa text encoder	80.4	12.1
Random text descriptions	78.5	14.0
Transformer EEG encoder only	74.2	18.3

Next, we measure zero-shot classification accuracy by not providing descriptions of target emotions during the supervised training on EEG data and applying prototype matching for test samples. In this setting, we obtain 25.3% zero-shot accuracy for DEAP arousal and 22.8% for valence. While these numbers lag behind the performance of fully supervised methods, they are noteworthy since there are no EEG samples of classes that need to be predicted.

The few-shot experiments reveal that the semantic space can be quickly adapted to by providing minimal number of labels. On the DEAP dataset with only 1% labeled EEG training samples, EEG-LM obtains 78.5% accuracy for arousal prediction, maintaining 84.9% of the baseline model's performance when fully labeled data is used. It is essential since gathering affective EEG data from the users in BCI systems is resource intensive and requires lots of effort. Also, the difference between zero and few-shot experiments indicates that the language prototype provides an essential prior, while only a little amount of labeled EEG data is still important for adaptation to the user-specific distribution of signals.

The ablation experiments confirm the complementarity of the proposed modules. Without the diffusion module, the model's performance drops from 92.5% to 88.2%. Removing the InfoNCE alignment objective causes a larger drop to 82.5%, indicating that semantic contrastive learning is the dominant contributor. Replacing emotion descriptions with random text reduces accuracy to 78.5%, further confirming that the gains come from meaningful affective language supervision rather than from additional model capacity alone.

Table 4 Generative quality comparison for EEG augmentation. Lower FID and higher Inception Score are better.

Augmentation Method	FID ↓	Inception Score ↑	DEAP Arousal Acc. (%)
No augmentation	–	–	82.5
GAN augmentation	28.4	2.1	66.8
DDPM augmentation	22.1	2.8	72.5
Latent diffusion	18.7	3.4	74.2
Conditional diffusion in EEG-LM	**15.3**	**4.1**	**92.5**

3.4 Interpretability and Generative Quality

t-SNE visualisation in Figure 2 shows that EEG-LM forms compact and well-separated arousal clusters after cross-modal alignment. Compared with unaligned EEG embeddings, the aligned representation reduces overlap between low- and high-arousal samples, suggesting that the language-guided objective reshapes the latent space into a more semantically ordered structure. This supports the quantitative improvements in Table 1 and explains why nearest-prototype zero-shot inference remains feasible.

From the SHAP summary presented in Figure 3 we can observe that the most predictive channels include fronto-central channels such as Fz, FCz, and Cz. Since these channels have been linked to functions such as attention and affective processing, it is evident that EEG-LM makes use of neurophysiological predictors rather than artifacts. The agreement between attention patterns and SHAP importance improves the interpretability of the model and makes its decisions easier to inspect in BCI and clinical settings.

We also assess the quality of generated EEG used for augmentation. The proposed conditional diffusion module obtains an FID of 15.3 and an Inception Score of 4.1, outperforming GAN augmentation (28.4/2.1) and standard DDPM augmentation (22.1/2.8). Lower FID indicates that generated samples are closer to real EEG distributions, while the higher Inception Score suggests greater diversity across affective conditions. Together with the ablation results, these findings show that diffusion improves generalisation by enriching the training distribution, while the largest discriminative benefit comes from aligning EEG features with meaningful language embeddings.

4 Conclusion

This study introduced EEG-LM, a foundation-style framework for EEG-based emotion recognition that connects neural signal representations with language-derived affective semantics. By combining a transformer EEG encoder, a frozen RoBERTa text encoder, InfoNCE-based cross-modal alignment, and conditional diffusion augmentation, the proposed model addresses three persistent challenges in affective EEG analysis: limited labelled data, poor cross-subject and cross-dataset generalisation, and weak semantic interpretability. The experimental results across DEAP, SEED, SEED-IV, and DREAMER demonstrate that language-grounded representation learning can

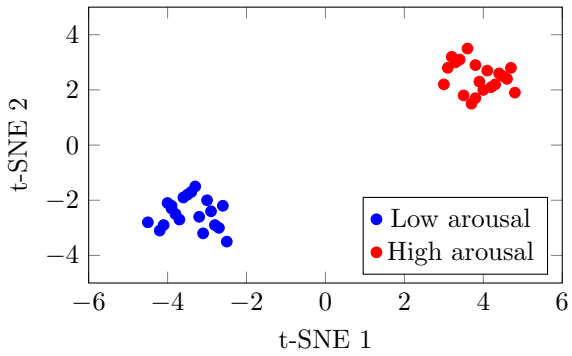


Fig. 2 t-SNE projection of EEG-LM embeddings for DEAP arousal classes.

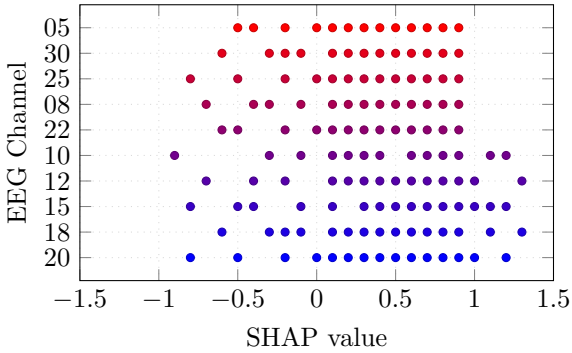


Fig. 3 SHAP summary plot with channels ordered by importance.

improve supervised recognition accuracy while also enabling zero-shot and few-shot inference through textual emotion prototypes.

Beyond improving classification performance, EEG-LM offers a more interpretable route for affective brain–computer interfaces. The shared EEG–text embedding space provides an explicit semantic target for neural representations, while the SHAP and attention analyses indicate that the model attends to plausible emotion-relevant EEG channels and patterns. These findings suggest that aligning brain signals with natural-language descriptions can make emotion decoding not only more robust, but also more transparent and easier to relate to human-understandable affect concepts.

A number of future directions seem to be essential. Firstly, further research should consider the use of larger and more diverse corpora of EEG recordings to test the consistency of the learned semantic space under different acquisition devices, cultures, stimulation protocols, and clinical populations. Secondly, future models may take into account other physiological features, like ECG, GSR, eye movements, and facial expressions, to build multimodal affective representations. Thirdly, distillation and compression techniques will have to be developed to implement real-time BCI applications on edge devices. And finally, clinical testing will require longitudinal controlled trials before EEG-LM can be used for monitoring mental health, neurofeedback, or decision-making assistance. In general, the findings show that EEG–LLM alignment

augmented with diffusion provides an important contribution towards data-efficient and generalisable brain emotion recognition models.

Credit authorship contribution statement

Praveen Goyal: Designed the study, performed the experiments, analyzed the data, prepared the figures, and wrote the manuscript. **Shyam Maheshwari** contributed to the conceptualization, validation, formal analysis, and manuscript revision. **Pankaj Pandey** supervised the study and critically reviewed the manuscript. All authors read and approved the final manuscript.

Funding

The authors received no funding from any organization for the submitted work.

Data Availability Statement

The datasets used and/or analysed during the current study include the DEAP, SEED, and DREAMER EEG datasets, together with synthetically generated data produced during the present study. The DEAP dataset is available from Queen Mary University of London at https://www.eecs.qmul.ac.uk/mmv/datasets/deap/. The SEED dataset is available from the BCMI Laboratory, Shanghai Jiao Tong University, at https://bcmi.sjtu.edu.cn/home/seed/, subject to the applicable access and usage requirements. The DREAMER dataset is available through the Zenodo repository at https://zenodo.org/records/546113 (DOI: https://doi.org/10.5281/zenodo.546113), subject to the applicable access conditions. The original third-party datasets are not redistributed by the authors. The synthetic data generated during the current study are available from the corresponding author on reasonable request.

Code Availability

The source code and implementation materials for EEG-LM are publicly available at https://github.com/praveensoit/EEG-LM.

Declaration of competing interest

We declare that the authors have no competing interests as defined by Nature Research, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

References

[1] A. Mishra, S. Shukla, J. Torres, J. Gwizdka, and S. Roychowdhury, "Thought2Text: Text generation from EEG signal using large language models (LLMs)," *arXiv preprint arXiv:2410.07507*, Oct. 2024.

[2] J. W. Kim, A. Alaa, and D. Bernardo, "EEG-GPT: Exploring capabilities of large language models for EEG classification and interpretation," *arXiv preprint arXiv:2401.18190*, 2024.

[3] J. Zhou, Y. Duan, F. Chang, T. Do, Y.-K. Wang, and C.-T. Lin, "BELT-2: Bootstrapping EEG-to-language representation alignment for multi-task brain decoding," *arXiv preprint arXiv:2409.00121*, 2024.

[4] H. Chen, W. Zeng, C. Chen, L. Cai, F. Wang, Y. Shi, L. Wang, W. Zhang, Y. Li, H. Yan, W. T. Siok, and N. Wang, "EEG emotion copilot: Optimizing lightweight LLMs for emotional EEG interpretation with assisted medical record generation," *Neural Networks*, vol. 192, p. 107848, 2025.

[5] Y. Duan, J. Zhou, Z. Wang, Y.-K. Wang, and C.-T. Lin, "DeWave: Discrete EEG waves encoding for brain dynamics to text translation," *arXiv preprint arXiv:2309.14030*, Jan. 2024.

[6] Z. Wang and H. Ji, "Open vocabulary electroencephalography-to-text decoding and zero-shot sentiment classification," *arXiv preprint arXiv:2112.02690*, Jan. 2024.

[7] T. C. N'dir and R. T. Schirrmeister, "EEG-CLIP: Learning EEG representations from natural language descriptions," *arXiv preprint arXiv:2503.16531*, 2025.

[8] W.-B. Jiang, L.-M. Zhao, and B.-L. Lu, "Large brain model for learning generic representations with tremendous EEG data in BCI," *arXiv preprint arXiv:2405.18765*, 2024.

[9] C.-S. Chen, Y.-J. Chen, and A. H.-W. Tsai, "Large cognition model: Towards pretrained EEG foundation model," *arXiv preprint arXiv:2502.17464*, 2025.

[10] G. Wang, W. Liu, Y. He, C. Xu, L. Ma, and H. Li, "EEGPT: Pretrained transformer for universal and reliable representation of EEG signals," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 37, 2024.

[11] X. Huang, C. Li, A. Liu, R. Qian, and X. Chen, "EEGDfus: A conditional diffusion model for fine-grained EEG denoising," *IEEE J. Biomed. Health Inform.*, 2024.

[12] J. Huang, M. Li, and W. Chen, "SAD-VER: A self-supervised, diffusion probabilistic model-based data augmentation framework for visual-stimulus EEG recognition," *Advanced Engineering Informatics*, vol. 65, p. 103298, 2025.

[13] S. Torma and L. Szegletes, "Generative modeling and augmentation of EEG signals using improved diffusion probabilistic models," *J. Neural Eng.*, vol. 22, no. 1, p. 016001, 2025.

[14] M. Li, P. Yu, and Y. Shen, "A spatial and temporal transformer-based EEG emotion recognition in VR environment," *Front. Hum. Neurosci.*, vol. 19, p. 1517273, 2025.

[15] H. Li and Y. Li, "Research on two-dimensional convolution-based EEG emotion recognition methods," in *Proc. 2024 Int. Conf. Biomedicine and Intelligent Technology*, 2024, pp. 147–151.

[16] Y. Li, Z. Kang, S. Gong, W. Dong, W. Zeng, H. Yan, W. T. Siok, and N. Wang, "Neural-MCRL: Neural multimodal contrastive representation learning for EEG-based visual decoding," in *2025 IEEE Int. Conf. Multimedia and Expo (ICME)*, 2025, pp. 1–6.

[17] Y. Du, Z. Guan, W. Huang, X. Zhang, and Q. Huang, "A case-based channel selection method for EEG emotion recognition using interpretable transformer networks," in *Proc. 2023 Int. Conf. Computer, Vision and Intelligent Technology (ICCVIT'23)*, 2023.

[18] Y. Li, S. Gong, W. Zeng, N. Wang, and W. T. Siok, "FreqDGT: Frequency-adaptive dynamic graph networks with transformer for cross-subject EEG emotion recognition," *arXiv preprint arXiv:2506.22807*, 2025.

[19] H. Wen and T. Yu, "A transformer framework based on self-supervised contrastive learning for EEG-based emotion recognition," in *2024 7th Int. Conf. Advanced Algorithms and Control Engineering (ICAACE)*, 2024, pp. 313–317.

[20] H. Wang and L. Song, "Supervised contrastive learning for EEG-based cross-subject emotion recognition," in *2023 3rd Int. Conf. Networking Systems of AI (INSAI)*, 2023, pp. 401–405.

[21] K. Patel, F. Safavi, R. Chandramouli, and R. Vinjamuri, "Transformer-based emotion recognition with EEG," in *2024 46th Annu. Int. Conf. IEEE Engineering in Medicine and Biology Society (EMBC)*, 2024, pp. 1–4.

[22] B. Mouazen, A. Benali, N. T. Chebchoub, E. H. Abdelwahed, and G. De Marco, "Enhancing EEG-based emotion detection with hybrid models: Insights from DEAP dataset applications," *Sensors*, vol. 25, no. 6, p. 1827, 2025.

[23] P. Patel, S. Balasubramanian, and R. N. Annavarapu, "Cross subject emotion identification from multichannel EEG sub-bands using Tsallis entropy feature and KNN classifier," *Brain Informatics*, vol. 11, 2024.

[24] N. Sharma and T. R. Bollu, "Explainable AI methods for interpreting emotions in brain–computer interface EEG data," in *Discovering the Frontiers of Human-Robot Interaction*, R. Vinjamuri, Ed. Cham: Springer, 2024.

[25] M. N. Imtiaz and N. Khan, "Enhanced cross-dataset electroencephalogram-based emotion recognition using unsupervised domain adaptation," *Comput. Biol. Med.*, vol. 184, p. 109394, Jan. 2025.

[26] Y.-D. Zhao, Y.-K. Liu, W.-L. Zheng, and B.-L. Lu, "EEG data augmentation for emotion recognition using diffusion model," in *2024 46th Annu. Int. Conf. IEEE Engineering in Medicine and Biology Society (EMBC)*, 2024, pp. 1–4.