Predicting Long-Term Movie Recall from EEG with Deep Models

Iván Martín-Fernández[1,*], Jaime León[1], Sergio Esteban-Romero[1], Manuel Gil-Martín[1] and Fernando Fernández-Martínez[1]

[1]*Grupo de Tecnología del Habla y Aprendizaje Automático (THAU), Information Processing and Telecommunications Center (IPTC), E.T.S.I. de Telecomunicación, Universidad Politécnica de Madrid (UPM), Madrid, 28040, Spain*

Abstract

This paper describes the THAU-UPM submissions to the MediaEval 2026 Predicting Movie and Commercial Memorability Task, focusing on EEG-based prediction of movie recall. We compare lightweight convolutional architectures, EEGNet, and pre-trained SignalJEPA representations under subject-independent and subject-dependent training regimes. Our results show that deep learning models can capture weak but meaningful signals from EEG, although performance remains below that of previous feature-based approaches. The best official test result is obtained with subject-dependent adaptation of pre-trained SignalJEPA features, reaching an AUC of 0.582 and suggesting that subject variability remains a major bottleneck for EEG-based recall prediction.

1. Introduction and Related Work

This paper reports the efforts of the THAU-UPM Team in the 2026 MediaEval Predicting Memorability Challenge [1]. In particular, we focus on Challenge 1.2: predicting subject-wise long-term movie recall using Electroencephalography (EEG) signals. The participants were given a set of EEG epochs recorded from different subjects while watching excerpts of movies and a binary self-reported annotation representing whether that subject recalls having watched said clip at any given point in their lives.

The data for this challenge was first used and released in the past edition of the benchmark, situating it in very early stages of development. Last year, the teams obtained modest but promising results, mainly using feature-driven machine learning approaches. The exception was the DACS-UM-RTL team, which achieved the second best score (0.635 AUC) by training a 1D Convolutional Neural Network (CNN) on the raw EEG data [2]. However, the best performance in the challenge (0.656 AUC) was obtained using a subject-dependent ElasticNet based system, in which each expert model was trained only with data belonging to the subject being evaluated [3].

Our approach for this edition is based on these learned lessons: the potential of deep learning architectures for the task and the importance of training recipes. We go beyond and explore different deep learning based architectures, either trained from scratch or pre-trained on large scale EEG data, and compare subject-independent training regimes with subject-dependent expert systems.

MediaEval'26: Multimedia Evaluation Workshop, June 15–16, 2026, Amsterdam, Netherlands and Online
*Corresponding author.
ivan.martinf@upm.es (I. Martín-Fernández)
0009-0004-2769-9752 (I. Martín-Fernández); 0009-0007-2835-2234 (J. León); 0009-0008-6336-7877 (S. Esteban-Romero); 0000-0002-4285-6224 (M. Gil-Martín); 0000-0003-3877-0089 (F. Fernández-Martínez)

2. Approach

Our proposal consists of a simple binary classification pipeline to predict long-term movie recall from EEG data using deep neural network backbones. We compare well-known architectures used for similar tasks in the literature with custom counterparts and explore the impact of using pre-trained checkpoints versus training the network from scratch. A detailed description of each of our runs is included below.

2.1. Run 1: Custom CNN

Inspired by the experiences reported by Ganesh et al. [2] and Huijben et al. [4], we design a 1D-CNN based pipeline for predicting movie recall. It consists of two convolutional layers, each followed by batch normalization, ReLU activation, max pooling, and 0.2 dropout. The final classification interface is a linear layer. We use this run to evaluate how traditional, task-based model design behaves in this setup.

2.2. Run 2: EEGNet

EEGNet is a well established, deliberately lightweight deep learning architecture designed for EEG-based Brain Computer Interfaces (BCIs) [5]. It mimics classical EEG processing from a data-driven perspective by automatically learning time-frequency selective filters, and then condensing information in the spatial domain. In contrast to Run 1, this approach exploits frequency and location-based relationships instead of relying only on temporal processing. It provides a strong baseline using a pre-defined and broadly explored architecture.

2.3. Run 3: Pre-trained SignalJEPA

In order to test the capabilities of EEG foundation models, pre-trained on large-scale general purpose data sources, we perform transfer learning on the SignalJEPA pre-trained model [6]. This architecture is based on the Joint Embedding Predictive Architecture (JEPA) [7], by which a modality encoder can be trained in a supervised manner to reconstruct masked tokens in the latent embedding space. We leverage this approach, designed to learn transferable representations from raw signals, to investigate whether generalistic features are useful for the recall prediction task.

In particular, we resort to the pre-local variant in which the signals are aggregated in the spatial dimension and fed into a CNN-based encoder and a fully connected layer that serves as a predictor. We keep the encoder frozen and train the prediction layer and spatial aggregation module to adapt them to the dimensions and characteristics of the dataset. The starting checkpoint[1] is pre-trained on BCI-related data from 40 subjects.

2.4. Run 4: Pre-trained SignalJEPA - Subject Dependent

Based on the hypothesis that deep learning approaches will be more data hungry than feature-based, but also able to generalize better across subjects, we train our models for runs 1-3 on all available data. However, motivated by the top result obtained in the 2025 edition [3] we submit a control run in which we train a specific system per subject, using the best-performing model from our LOSO validation experiments. The rest of the configuration is identical to Run 3.

[1] https://huggingface.co/braindecode/signal-jepa_without-chans

Table 1
Results on the development set. **Boldface** indicates best run per metric.

Run	Macro acc.	Balanced acc.	F1	AUC
1 - Custom CNN	0.520	0.512	0.373	**0.524**
2 - EEGNet	0.500	0.504	**0.432**	0.516
3 - Pre-trained SignalJEPA	**0.539**	**0.514**	0.384	0.519
4 - Pre-trained SignalJEPA - Subject Dependent	0.520	0.505	0.308	0.521

2.5. Experimental Setup

Regarding signal pre-processing, we take the 0 - 4 s window for each epoch, resample it to 64 Hz, and perform z-score channel-wise normalization. No other pre-processing stage is applied in order to defer pattern analysis and feature identification to the deep learning models.

Our approaches are validated on the development set using Leave One Subject Out Cross Validation (LOSO CV) to evaluate their generalization, with the exception of run 4 which was evaluated using an in-house 70% / 15% / 15% train-val-test split per subject. We emphasize the use of LOSO CV, as it is key for developing systems that are deployable in the wild, where no previous data is available for a new subject. In addition to the Area Under the Receiver-Operating Characteristic Curve (AUC), the official challenge metric, we also report macro and balanced accuracy, as well as the F1 Score on the development set. This is done to provide additional analysis of system performance and to identify pitfalls and failure modes. In order to generate binary predictions, a threshold of 0.5 is selected for all systems. No calibration was performed. All experiments were performed on a MacBook Pro M3 with 36 GB of unified memory.

3. Results and Analysis

3.1. Development Results

Table 1 shows the results of our runs in our in-house development setup. It is worth noting that the metrics of Run 4 are not directly comparable to the rest of the runs, as the evaluation setup shifts from LOSO to a per-subject held-out setting.

When comparing across model architectures, none of them achieves the best performance across all evaluation metrics. Moreover, the differences across approaches are generally modest. Nevertheless, some interpretable patterns emerge.

The pre-trained SignalJEPA obtains the best macro and balanced accuracy, suggesting slightly better overall class-wise classification performance. However, the margin is small, particularly in balanced accuracy, where its performance is very close to that of the custom CNN. EEGNet obtains the best F1 score, suggesting a more favorable precision-recall trade-off for the positive class under the selected decision threshold. Finally, the custom CNN obtains the best AUC, indicating that it provides the strongest threshold-independent ranking of positive and negative samples, even if this does not translate into the best thresholded classification performance.

Regarding Run 4, although its results are not directly comparable with the rest of the systems, performance remains roughly in the same range. However, its lower F1 score suggests that the subject-dependent setup does not clearly improve positive-class detection in this configuration. Overall, the differences across metrics do not reveal a clear preference for model selection on the development set.

Table 2
Results on the testing set. PT refers to Pre-trained. **Boldface** indicates best AUC.

Run	# Total Param.	# Trained Param.	Dev AUC	Test AUC
1 - CNN	17,826	17,826	**0.524**	0.555
2 - EEGNet	1,874	1,874	0.516	0.545
3 - PT SignalJEPA	14,486	646	0.519	0.503
4 - PT SignalJEPA - Subject Dep.	14,486	646	0.521	**0.582**

3.2. Official Test Results

Table 2 shows the results of our runs on the official challenge testing split in terms of AUC, the official challenge metric. We also report the total and trainable parameter counts as indicators of model size and training budget.

All runs remain relatively close to chance, underscoring the difficulty of the task, as already observed in previous editions. Moreover, our approaches remain below the best-performing feature-based systems, suggesting that task-specific knowledge and carefully designed representations remain highly relevant.

When comparing runs, the best test performance is obtained by the subject-dependent SignalJEPA-based system, suggesting that subject-specific modeling is beneficial, even when trained with less data than the subject-independent alternatives. This is consistent with the previous edition [3], and provides further evidence that subject variability and generalization remain among the main bottlenecks of the task. Differences across architectures remain modest, and ordering is not preserved. The custom CNN remains the best-performing system after excluding Run 4, while EEGNet surpasses SignalJEPA, the only system whose performance drops from development to testing. This suggests that the pre-trained representation does not necessarily transfer robustly under the subject-independent setting.

Regarding model size, EEGNet captures a similar amount of predictive signal to the custom CNN while using roughly 10% of its parameters. This supports the importance of architectural design and EEG-specific inductive biases, rather than simply increasing parameter count. At the same time, the best test result is obtained by training only 646 parameters, suggesting that robust pre-trained representations may still be useful when combined with subject-specific adaptation.

4. Discussion and Outlook

We have presented the contributions of THAU-UPM to the MediaEval Predicting Movie and Commercial Memorability Task, specifically to EEG-based prediction of movie recall. Our comparison of different deep learning architectures, pre-training regimes, and data recipes revealed that lightweight convolutional models can capture meaningful signal for the task, but do not surpass previous feature-based approaches. This points towards the need for a better understanding of the task and data to inform model design, and highlights the limitations of end-to-end deep learning approaches when expert knowledge is not explicitly incorporated.

Even though model design, architecture size, and pre-training data influence performance, the most relevant factor remains the task-specific training regime. A central challenge is, therefore, to design models that capture meaningful intra- and inter-subject representations. We hope this research drives future model design towards architectures that combine data-driven learning with expert knowledge to achieve accurate, subject-independent models of recall.

Acknowledgements

The research of Iván Martín-Fernández was supported by the Universidad Politécnica de Madrid (Programa Propio I+D+i). This work was funded by the Spanish Ministry of Science and Innovation through the project TRUSTBOOST (PID2023-150584OB-C21), funded by MCIN/AEI/10.13039/501100011033 and by the European Union "NextGenerationEU/PRTR".

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT in order to: Improve writing style, Grammar and spelling check. After using these tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] I. Martín-Fernández, A. Ganesh, M. G. Constantin, C.-H. Demarty, M. Gil-Martín, S. Halder, B. Ionescu, A. Matran-Fernandez, R. Savran Kiziltepe, A. García Seco de Herrera, Overview of the mediaeval 2026 predicting movie and commercial memorability task, in: Proc. of the MediaEval 2026 Workshop, Amsterdam, The Netherlands and Online, 2026. In Press.

[2] A. Ganesh, I. Huijben, B. Khaertdinov, J. Iskaj, M. Popa, N. Tintarev, DACS-UM-RTL: Early fusion and pre-text task learning for video memorability prediction, in: Working Notes Proceedings of the MediaEval 2025 Workshop, Dublin, Ireland and Online, 25-26 October 2025, 2025. In Press.

[3] I. Martín-Fernández, M. Lobo-Alonso, S. Esteban-Romero, M. Gil-Martín, F. Fernández-Martínez, Exploring Movie Recall Prediction Using Functional Descriptors of the EEG Signal, in: Working Notes Proceedings of the MediaEval 2025 Workshop, Dublin, Ireland and Online, 25-26 October 2025, 2025. In Press.

[4] I. Huijben, A. Ganesh, A. Wodeyar, P. Bonizzi, Subject Variability and Spatio-Temporal Dynamics in EEG During Video Recall, in: embc, IEEE, 2026. In Press.

[5] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, B. J. Lance, EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces, Journal of Neural Engineering 15 (2018) 056013. doi:10.1088/1741-2552/aace8c.

[6] Guetschel, Pierre, Moreau, Thomas, Tangermann, Michael, S-JEPA: TOWARDS SEAMLESS CROSS-DATASET TRANSFER THROUGH DYNAMIC SPATIAL ATTENTION, in: Proceedings of the 9th Graz Brain-Computer Interface Conference 2024, Verlag der Technischen Universität Graz, 2024, pp. 11–16. doi:10.3217/978-3-99161-014-4-003.

[7] Y. LeCun, A Path Towards Autonomous Machine Intelligence Version 0.9.2, 2022-06-27, 2022. URL: https://openreview.net/pdf?id=BZ5a1r-kVsf.