MediaEval 2025 : A Multimodal Approach for Predicting Movie and Commercial Memorability using Stacking and Gradient Boosting

Sakthi Mukesh Thanga Mariappan[1,*,†], Muthulakshmi Ramasamy[2,†] and Beulah Arul[1,†]

[1] *Rajalakshmi Engineering College, Chennai, Tamil Nadu, India*
[2] *Francis Xavier Engineering College, Tirunelveli, Tamil Nadu, India*

Abstract

This paper details our participation, team "CodingSoft-REC", in the MediaEval 2025 "Predicting Media Memorability" challenge, addressing four subtasks across two new datasets for movies and commercials. For movie memorability, we employed a stacked ensemble of Ridge regressors using ResNet50, R3D, and ViT features, achieving a Spearman correlation of 0.224. For EEG-based recall prediction, an XGBoost classifier yielded an AUC of 0.511, highlighting the difficulty of the task. In the commercial video domain, we used a multimodal XGBoost model that combined visual features (AlexNet, DenseNet, etc.) with engagement metadata to predict video and brand memorability. This approach proved effective for predicting video recall (Spearman = 0.231) but struggled with brand recall (Spearman = -0.075), suggesting that distinct features are needed for each prediction target. Our results demonstrate the utility of ensemble methods and underscore the challenge of modeling brand memorability and interpreting neural correlates of memory.

1. Introduction

Predicting which videos and scenes will be remembered has significant applications across media studies, advertising, and cognitive neuroscience. The MediaEval 2025 Predicting Media Memorability task [1] introduced two novel datasets—one focusing on movies and one on commercials—challenging participants to build models for automatic memorability prediction and to analyze which features best explain it.Understanding these features is crucial, as memorability is a key factor in the viral potential of online content and the effectiveness of marketing campaigns.

Uniquely, the 2025 edition also integrates EEG data [2] for probing human recall and introduces the new challenge of predicting brand recall from advertisements. This work documents our participation across all four subtasks, where we applied established machine learning techniques to these novel, multimodal datasets. We utilized the provided features and analyzed the effectiveness of stacking [3] and boosting models [4] for these memorability prediction tasks.

MediaEval'25: Multimedia Evaluation Workshop, October 25–26, 2025, Dublin, Ireland and Online

[*] Corresponding author.

[†] These authors contributed equally.

sakthiimukesh@gmail.com (S. M. T. Mariappan); muthulakshmiiiir@gmail.com (M. Ramasamy); beulah.a@rajalakshmi.edu.in (B. Arul)
0009-0005-3044-6841 (S. M. T. Mariappan); 0009-0009-9911-5812 (M. Ramasamy); 0000-0002-3891-0806 (B. Arul)

CEUR Workshop Proceedings (CEUR-WS.org)

2. Related Work

Prior research in media memorability ranges from classic cognitive studies to recent deep learning approaches that utilize audio-visual cues, text, and engagement data. Previous MediaEval editions have demonstrated the value of multimodal features and large-scale annotation in modeling video memorability. The work by Cohendet et al. [5] (2018) introduced long-term video memorability datasets and analysis techniques, providing a foundational methodology for this task. Furthermore, the use of EEG for memorability detection remains an emerging field of investigation, as proposed by Matran-Fernandez and Halder (2025) [6]. Our approach builds on these foundations and adapts literature-proven regression and classification models to the expanded datasets and new tasks of MediaEval 2025.

3. Approach

Our strategy for the MediaEval 2025 memorability tasks was centered on robust ensemble learning techniques tailored to the specific challenges and data modalities of each subtask.

3.1. Subtask 1: Movie Memorability (Video and EEG)

This subtask is for predicting movie memorability scores [7](Subtask 1.1) and detecting recall from EEG data (Subtask 1.2). We thought of training it with distinct models would be better for their respective use cases and their complexity.

3.1.1. Subtask 1.1 (Video Memorability)

For predicting movie memorability scores (Subtask 1.1) and detecting recall from EEG data (Subtask 1.2), we employed distinct models suited for each data type.

We utilized a selection of the provided pre-extracted visual features, focusing on ResNet50 [8], R3D [9], and ViT-B16 [10]. For frame-based features, we aggregated their vectors using mean pooling [11] to create a single representative vector per video.

Our architecture consisted of a two-level stack. Three base Ridge regressors [12] were trained independently on each feature set. A final "meta" Ridge regressor then combined the predictions from these base models to produce the final memorability score.

Models were trained within a pipeline that included StandardScaler for normalization. We performed a cross-validation grid search with k-fold method to find the optimal alpha (regularization strength) for each Ridge model, optimizing for the Spearman correlation coefficient.

3.1.2. Subtask 1.2 (EEG-based Recall)

For this classification task, we used an XGBoost model. The approach from Subtask 1.1 was adapted to handle the provided EEG features [13] (ERP and ERSPs) are taken as the exact value without any pre-processing or normalization,to predict whether a participant remembered a video. This allowed us to apply a powerful gradient boosting method directly to the neural data.

3.2. Subtask 2: Commercial & Brand Memorability

For predicting both commercial video memorability [14] (Subtask 2.1) and brand memorability (Subtask 2.2), we adopted a multimodal approach using a powerful gradient boosting framework for the VIDEM dataset [15].

Regarding feature engineering, We created a comprehensive feature set by combining both visual and tabular data.

- For **visual features**, we horizontally stacked the flattened feature vectors from *AlexNet*, *DenseNet121*, *EfficientNetB3*, and *ResNet50*.
- For **tabular features**, we incorporated metadata such as `channelName` (label-encoded), `categoryId`, and `engagementRate` to capture non-visual cues.

XGBoost Model: We utilized an XGBoost [16] (Extreme Gradient Boosting) regressor to model the relationship between our comprehensive feature set and the memorability scores. XGBoost is an advanced implementation of gradient boosting that builds an ensemble of decision trees sequentially, with each new tree correcting the errors of the previous ones.

Training and Optimization: The XGBoost model was configured to use GPU acceleration for faster training. We tuned key hyperparameters, including the number of estimators, max tree depth, and learning rate, to optimize performance and prevent overfitting, using a 90/10 train-test split for validation.

4. Results and Analysis

Below, we present our official test results for the four MediaEval 2025 challenges. Performance was evaluated using Spearman's Rank Correlation Coefficient [17] (SRCC) and Mean Squared Error [18] (MSE) for regression tasks, and Area Under the ROC Curve [19] (AUC) for the EEG-based classification task.

Table 1
Official Test Results for MediaEval 2025 - Media Memorability Task

Challenge	Run ID	Spearman	Pearson	MSE
1.1	runStackModelRidge	0.224	0.190	0.102
1.2	runXgBoost	AUC: 0.511		—
2.1	runStackXG	0.231	0.247	0.025
2.2	runStackXGBoost	-0.075	-0.060	0.028

4.1. Analysis and Observations

Our stacking-based regression (Ridge) model for challenge 1.1 achieved a moderate correlation, confirming the utility of combining diverse video-level features. The positive Spearman score indicates our model successfully ranked the memorability of movie clips better than random chance.

The XGBoost model in challenge 1.2 ("EEG recall") achieved an AUC of 0.511, indicating a slight improvement over a random classifier, but highlighting the inherent challenge in using raw EEG data for subject recall. Room for improvement exists in advanced EEG preprocessing and deep sequence modeling.

For commercials (challenge 2.1), our StackXG model outperformed the movie regression setup, suggesting engagement and metadata features provide additional predictive power not available in the movie dataset.

Challenge 2.2 (brand recall) proved the hardest: a negative Spearman result points to overfitting or insufficient feature extraction relating to the brand, possibly due to the tight coupling of brand perception and creative context. This suggests that the features predictive of video memorability are not necessarily the same as those predictive of brand memorability.

5. Conclusion

Our participation in the MediaEval 2025 Predicting Media Memorability task demonstrated the importance of diverse visual and metadata features for video memorability prediction, and identified specific limitations when moving to subject-specific recall or brand memorability. While conventional stacking and boosting approaches remain competitive, future research should address model interpretability and specialized processing for EEG and engagement signals. Provided datasets offer unique opportunities to advance both computational modeling and understanding of human memory processes in multimedia contexts.

Declaration on Generative AI

During the preparation of this work, the author used generative AI tools solely for grammar and spelling checks. All content—including approach, analysis, and discussion—has been prepared and critically reviewed by the author to ensure originality and clarity.

References

[1] I. Martín-Fernández, 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 2025 predicting movie and commercial memorability task, in: Proc. of the MediaEval 2025 Workshop, Dublin, Ireland and Online, 2025.

[2] M. G. Constantin, B. Ionescu, C.-H. Demarty, N. Q. Duong, X. Alameda-Pineda, M. Sjöberg, The predicting media memorability task at mediaeval 2019., in: MediaEval, 2019.

[3] B. Pavlyshenko, Using stacking approaches for machine learning models, in: 2018 IEEE second international conference on data stream mining & processing (DSMP), IEEE, 2018, pp. 255–258.

[4] D.-S. Cao, Q.-S. Xu, Y.-Z. Liang, L.-X. Zhang, H.-D. Li, The boosting: A new idea of building models, Chemometrics and Intelligent Laboratory Systems 100 (2010) 1–11.

[5] R. Cohendet, K. Yadati, N. Q. Duong, C.-H. Demarty, Annotating, understanding, and predicting long-term video memorability, in: Proceedings of the ICMR 2018 Conference, Yokohama, Japan, 2018, pp. 11–14.

[6] A. Matran-Fernandez, S. Halder, An eeg dataset to study neural correlates of audiovisual long-term memory retrieval, Research Square (2025). doi:10.21203/rs.3.rs-7066609/v1, preprint.

[7] Y. Baveye, R. Cohendet, M. Perreira Da Silva, P. Le Callet, Deep learning for image memorability prediction: The emotional bias, in: Proceedings of the 24th ACM international conference on Multimedia, 2016, pp. 491–495.

[8] B. Koonce, Resnet 50, in: Convolutional neural networks with swift for tensorflow: image recognition and dataset categorization, Springer, 2021, pp. 63–72.

[9] Y. Sun, H. He, C. Tang, S. Huang, B. Wang, T. Jiang, Dynamic gesture recognition using r3d network with adaptive temporal feature resolutions, in: International Conference on Guidance, Navigation and Control, Springer, 2024, pp. 332–341.

[10] M. G. Dahmani, M. Tarhouni, S. Zidi, Vision transformers (vit) for enhanced skin cancer classification, in: 2024 IEEE International Conference on Artificial Intelligence & Green Energy (ICAIGE), IEEE, 2024, pp. 1–6.

[11] Y. Gu, C. Li, J. Xie, Attention-aware generalized mean pooling for image retrieval, arXiv preprint arXiv:1811.00202 (2018).

[12] G. C. McDonald, Ridge regression, Wiley Interdisciplinary Reviews: Computational Statistics 1 (2009) 93–100.

[13] A. Uran, C. Van Gemeren, R. van Diepen, R. Chavarriaga, J. d. R. Millán, Applying transfer learning to deep learned models for eeg analysis, arXiv preprint arXiv:1907.01332 (2019).

[14] S. Harini, S. Singh, Y. K. Singla, A. Bhattacharyya, V. Baths, C. Chen, R. R. Shah, B. Krishnamurthy,

Long-term ad memorability: Understanding & generating memorable ads, in: 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), IEEE, 2025, pp. 5707–5718.

[15] R. S. Kiziltepe, S. Sahab, R. V. Santana, F. Doctor, K. Paterson, D. Hunstone, A. G. S. de Herrera, VIDEM: VIDeo Effectiveness and Memorability Dataset, in: Proceedings of the 18th International Work-Conference on Artificial Neural Networks (IWANN 2025), A Coruña, Spain, 2025, pp. 16–18.

[16] J. Chen, F. Zhao, Y. Sun, Y. Yin, Improved xgboost model based on genetic algorithm, International Journal of Computer Applications in Technology 62 (2020) 240–245.

[17] P. Sedgwick, Spearman's rank correlation coefficient, Bmj 349 (2014).

[18] T. O. Hodson, T. M. Over, S. S. Foks, Mean squared error, deconstructed, Journal of Advances in Modeling Earth Systems 13 (2021) e2021MS002681.

[19] A. P. Bradley, The use of the area under the roc curve in the evaluation of machine learning algorithms, Pattern recognition 30 (1997) 1145–1159.