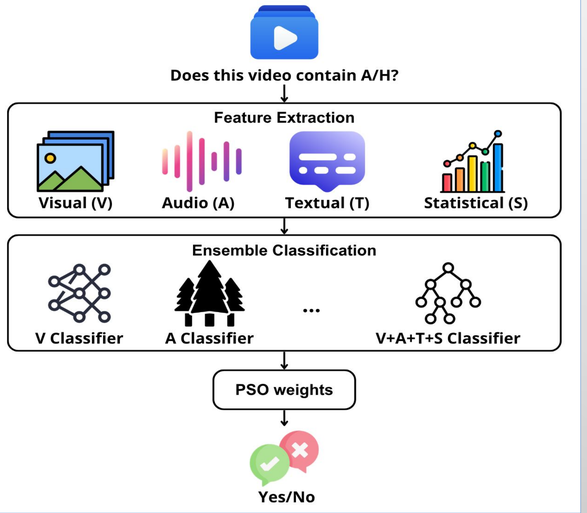
BROTHER: Behavioral Recognition Optimized Through Heterogeneous Ensemble Regularization for Ambivalence and Hesitancy

Alexandre Pereira, MD; Pablo Barros, PhD; Bruno Fernandes, PhD
University of Pernambuco

CVPR JUNE 3-7, 2026 DENVER COLORADO

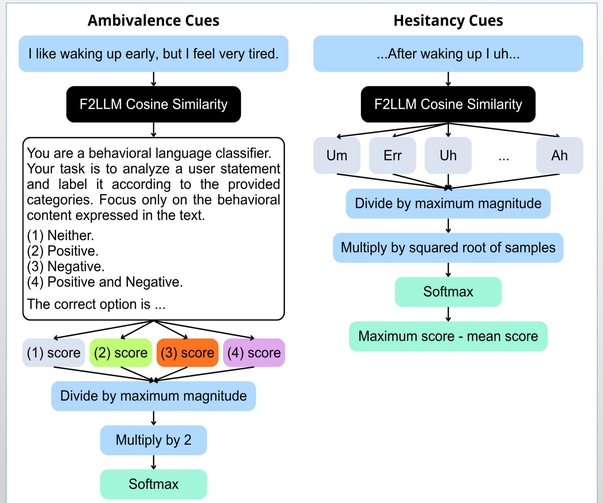
Abstract

- **Ambivalence and hesitancy (A/H)** are complex psychological states marked by conflicting behaviors, making them difficult to detect in real-world settings.
- Current multimodal systems often overfit constrained datasets and lose important cues during feature fusion.
- We propose **BROTHER**, a multimodal ensemble that combines visual, audio, textual, and statistical cues while using PSO to silence overfitted models.

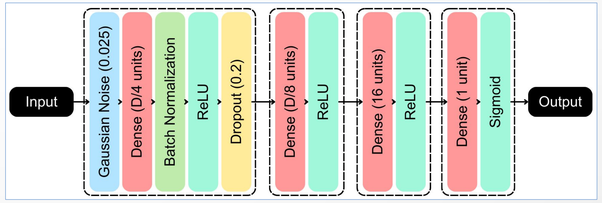


Methods and Materials

- **Visual features:** The person's face is cropped and encoded with **SigLip2**, followed by frame filtering, pooling, and PCA reduction.
- **Audio features:** Extracted with **HuBERT**, a non-supervised embedding model.
- **Textual features:** Encoded using **F2LLM**, a model trained exclusively on non-synthetic data.
- **Statistical features:** Analytics from the three previous modalities, including **frame similarity**, **audio features**, **textual patterns** and prompting for **A/H cues**.



- For each of the 15 modality combinations, three candidate architectures are trained: **Multi-Layer Perceptron**, **Random Forest**, and **Gradient Boosting Decision Tree**. The best architecture for each subset is selected by validation loss.



- The chosen models are fused using PSO-based hard voting. Each particle defines the influence of a model towards the final decision.

$$Fitness = \frac{2 \cdot F1_{val} \cdot F1_{train}}{F1_{val} + F1_{train}} - (\lambda \cdot |F1_{train} - F1_{val}|)^2$$

Results

- On the official hidden final test set, **BROTHER** achieved **0.7266 Macro F1**, obtaining **1st place in the challenge**.
- Text was the strongest standalone modality, while the Statistical modality improved stability and strengthened weaker visual and audio combinations.

Team	Val	Test	Final Test
VisPBF (Ours)	0.7355	0.7465	**0.7266**
Fennec	0.7460	0.6940	0.7151
LEYA	0.8300	0.8077	0.7142
Lenovo PCIE	–	–	0.6748
Time Visão	0.6524	0.6808	0.5362

Discussion

- Our findings suggest that A/H recognition benefits from treating **behavioral cues** as **complementary** and **unevenly distributed** across modalities.
- Instead of forcing all features into one model, the ensemble allowed **stronger combinations** to **contribute more** while limiting the effect of overfitted ones.

Conclusions

- BROTHER presents a framework for **A/H detection** in **real-world videos** by combining foundation model embeddings with a dedicated statistical modality.
- Its performance in the challenge highlights the value of designing A/H recognition systems around subtle behavioral conflict, rather than treating it as a standard emotion classification task.

SCAN ME

Contact

Alexandre Rodolfo Moreira Pereira
University of Pernambuco
Email: armp@ecomp.poli.br

Website: https://www.linkedin.com/in/alexandre-rodolfo-079433270/
Phone: +55 81 98819-5656

References

1. Manuela González-González et al. BAH dataset for ambivalence/hesitancy recognition in videos for digital behavioural change. In ICLR, 2026.
2. Andrey V. Savchenko. HSEmotion team at ABAW-8 competition: Audiovisual ambivalence/hesitancy, emotional mimicry intensity and facial expression recognition. In CVPRW, 2025.
3. Michael Tschannen et al. SigLIP 2: Multilingual vision language encoders with improved semantic understanding. arXiv preprint arXiv:2502.14786, 2025.
4. Wei-Ning Hsu et al. HuBERT: Self-supervised speech representation learning by masked prediction of hidden units. IEEE/ACM Trans. Audio, Speech, Lang. Process., 29:3451–3460, 2021.
5. Ziyin Zhang et al. F2LLM technical report: Matching SOTA embedding performance with 6 million open-source data. arXiv preprint arXiv:2510.02294, 2025.
6. Shayan Ali Bargal et al. Emotion recognition in the wild from videos using images. In Proc. ACM Int. Conf. Multimodal Interaction (ICMI), pages 433–436, 2016.