

REPORT

Multimodal emotion recognition leverages complementary information from audio, visual, and textual modalities to improve accuracy and robustness. [1] proposed TA-AVN, an end-to-end framework that uses temporal aggregation of asynchronous audio-visual data, achieving state-of-the-art results on CREMA-D and RAVDESS datasets while enabling real-time processing and data augmentation through random segment sampling [1]. [2] integrated EEG signals and facial expressions using CNNs and attention mechanisms, reporting high accuracy on valence and arousal dimensions in the DEAP and MAHNOB-HCI datasets . [3] introduced Af-CAN, a context-aware model utilizing Bi-GRU and attention mechanisms for conversational emotion recognition, showing superior performance on IEMOCAP and MELD by modeling speaker-specific and global contextual features . [4] addressed data scarcity via cross-modal translation using GANs and sequence-level discriminators, enabling the use of heterogeneous datasets without strict modality alignment and improving performance on CMU-MOSEI and IEMOCAP[4] . Together, these works highlight the importance of temporal modeling, contextual awareness, attention mechanisms, and cross-modal synthesis in advancing multimodal emotion recognition.

- [1] A. Radoi, A. Birhala, N. C. Ristea, and L. C. Dutu, "An End-To-End Emotion Recognition Framework Based on Temporal Aggregation of Multimodal Information," *IEEE Access*, vol. 9, pp. 135559–135570, 2021, doi: 10.1109/ACCESS.2021.3116530.
- [2] S. Wang, J. Qu, Y. Zhang, and Y. Zhang, "Multimodal Emotion Recognition From EEG Signals and Facial Expressions," *IEEE Access*, vol. 11, pp. 33061–33068, 2023, doi: 10.1109/ACCESS.2023.3263670.
- [3] X. Zhang, M. Wang, X. Zeng, and X. Zhuang, "Af-CAN: Multimodal emotion recognition method based on situational attention mechanism," *IEEE Access*, 2024, doi: 10.1109/ACCESS.2024.3471613.
- [4] Y. C. Yoon, "Can We Exploit All Datasets? Multimodal Emotion Recognition Using Cross-Modal Translation," *IEEE Access*, vol. 10, pp. 64516–64524, 2022, doi: 10.1109/ACCESS.2022.3183587.