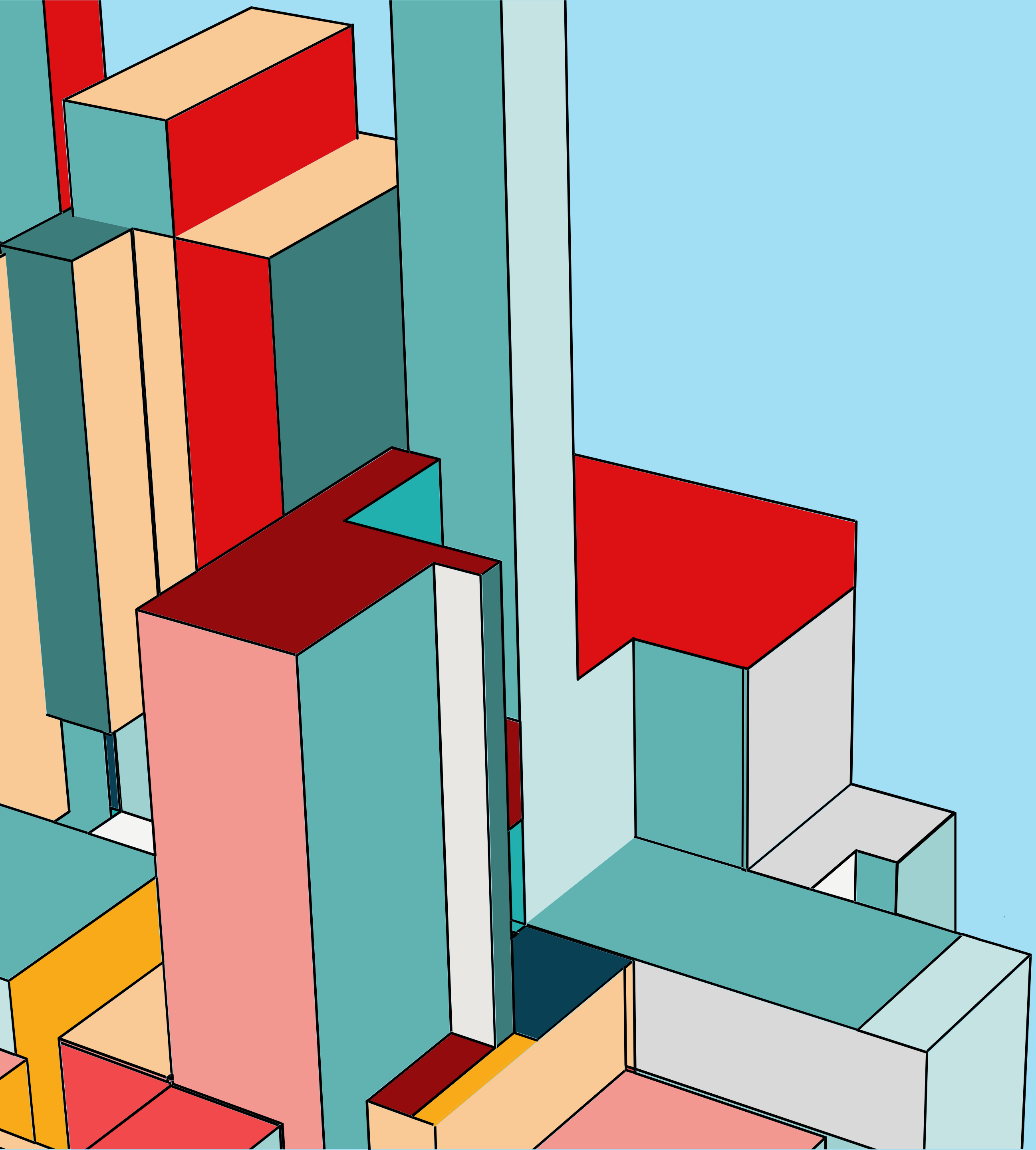


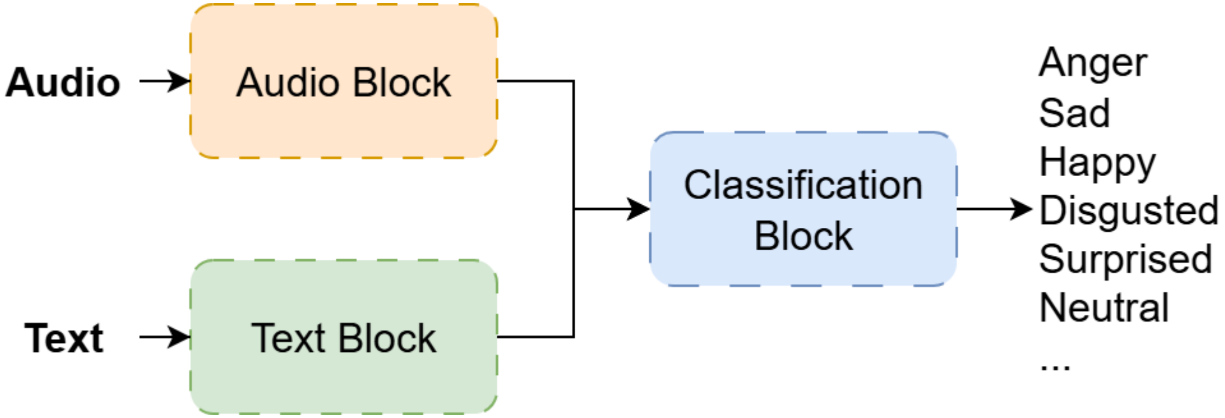
APPLIED DATA SCIENCE PROJECT

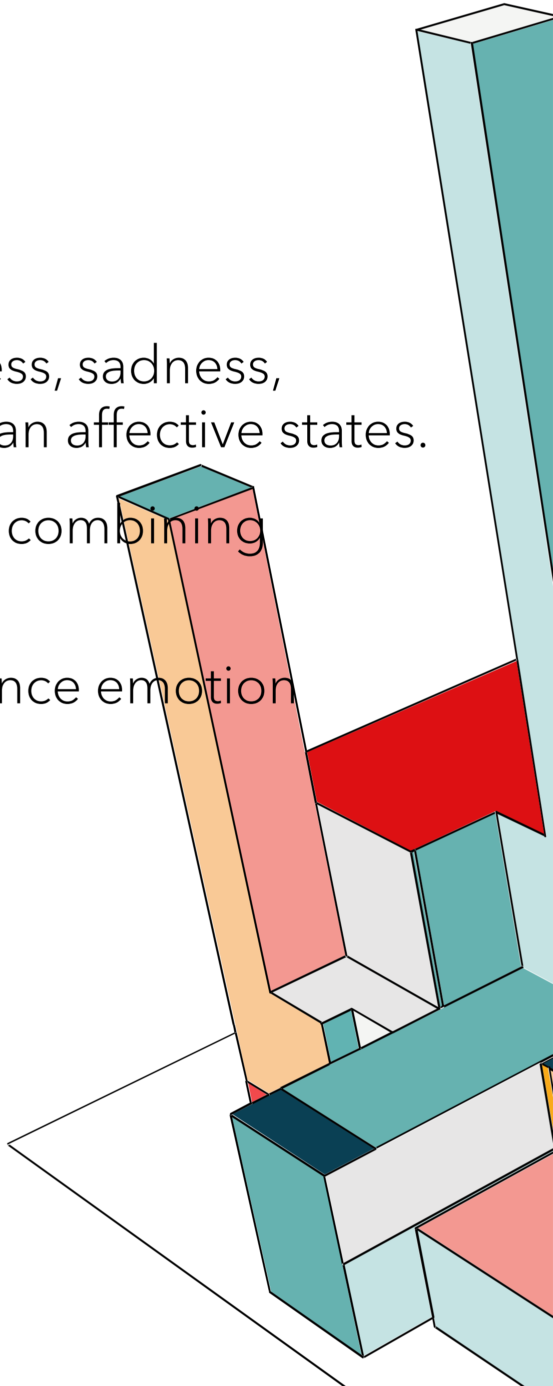
MULTIMODAL EMOTION RECOGNITION

1859
Politecnico di Torino
FONDAZIONE links PASSION FOR INNOVATION
ellis European Laboratory for Learning and Intelligent Systems

VALUE-DRIVEN PROJECT

- Emotion Recognition is the task to classify emotional states (e.g., happiness, sadness, anger, fear) so that machines can better understand and respond to human affective states.
- Multimodal emotion recognition is the process of detecting emotions by combining multiple data sources, such as speech and text.
- The goal of the project is to integrate speech and text modalities to enhance emotion recognition performance beyond using text or speech alone.

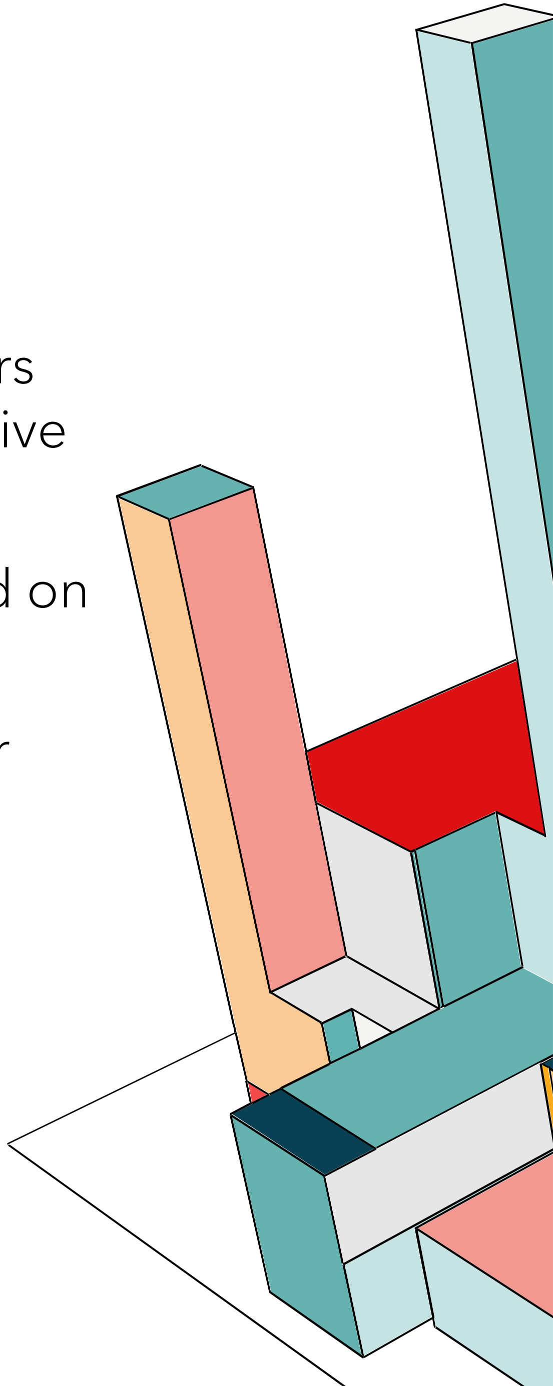




IMPACT AND USE CASES OF MULTIMODAL EMOTION RECOGNITION

- Support for Autism Care Centres: Emotion recognition can help caregivers detect early signs of emotional dysregulation in patients, enabling proactive intervention to prevent potential crises.
- Customer Experience: Helps businesses adapt services in real time based on user sentiment.
- Safety & Security: Detecting emotions in critical environments (e.g., driver fatigue, public spaces) can prevent accidents.
- Media & Entertainment: Personalisation of content based on viewers' emotional responses.

3 GOOD HEALTH AND WELL-BEING

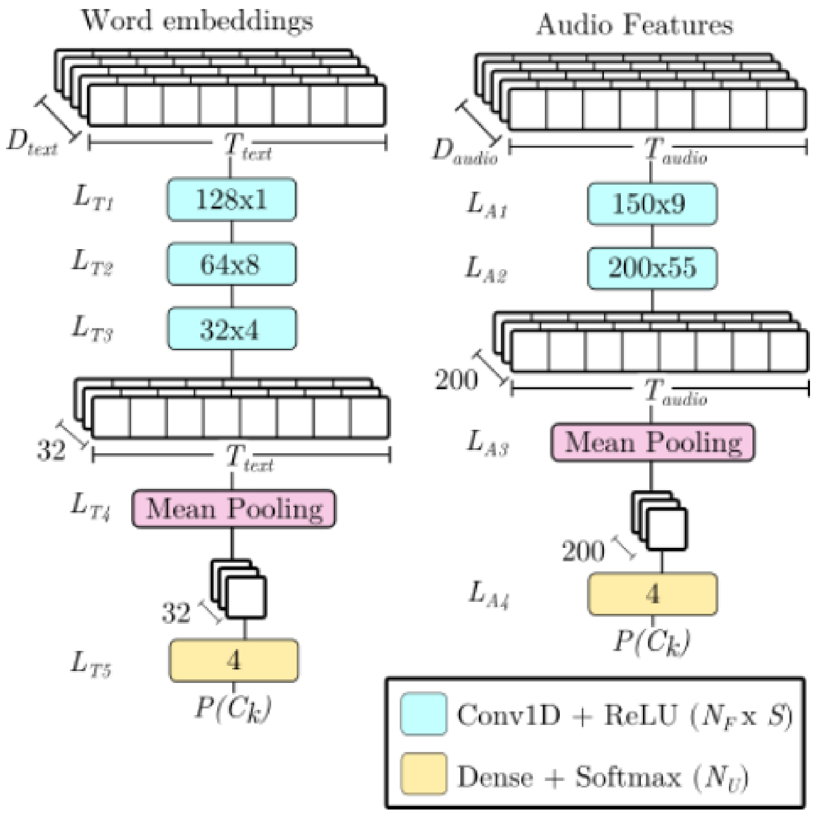


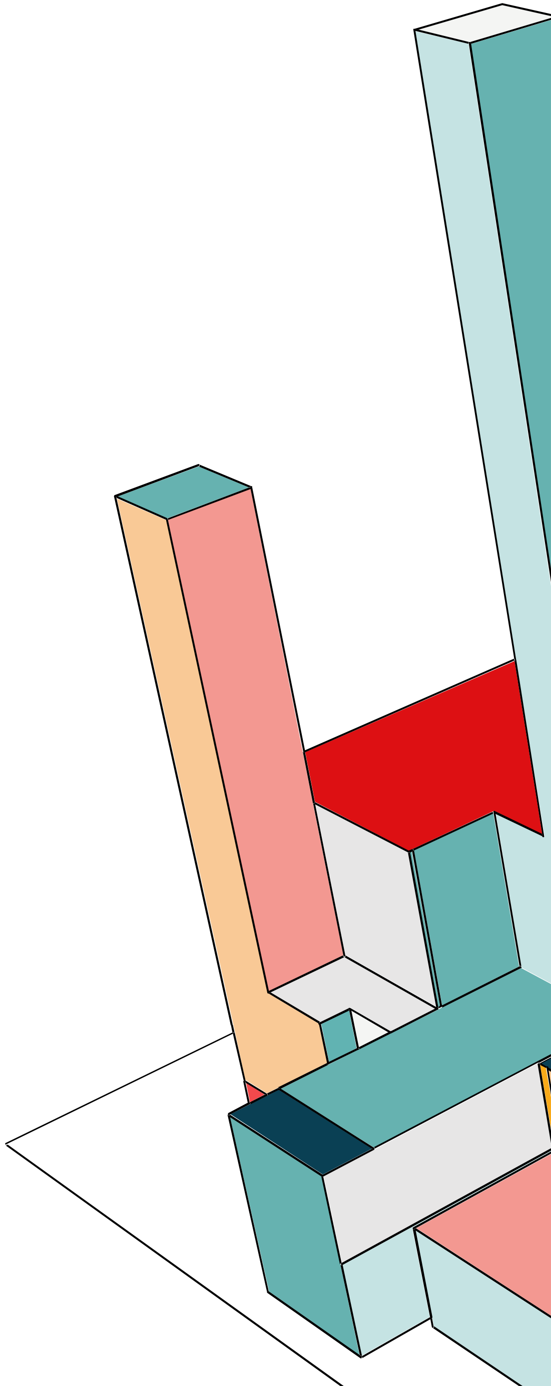
DATA

- IEMOCAP: IEMOCAP is a ~12-hour multimodal database (video, speech, motion capture, text) of dyadic acted scenarios, annotated with categorical and dimensional emotion labels, widely used for multimodal emotion research.
- MSP-Podcast: A large multimodal dataset of 100k+ English podcast episodes with audio, transcripts, and metadata, supporting tasks like summarization, retrieval, and emotion recognition.

TASK

- The goal is to build a text+audio emotion recognition model that outperforms audio-only or text-only models.
- First, replicate the architecture from Fusion Approaches for Emotion Recognition from Speech Using Acoustic and Text-Based Features.
- Then, improve it by using stronger audio models (e.g., Whisper) and text models (e.g., RoBERTa as alternative to BERT) or Audio-Text models (e.g., CLAP).





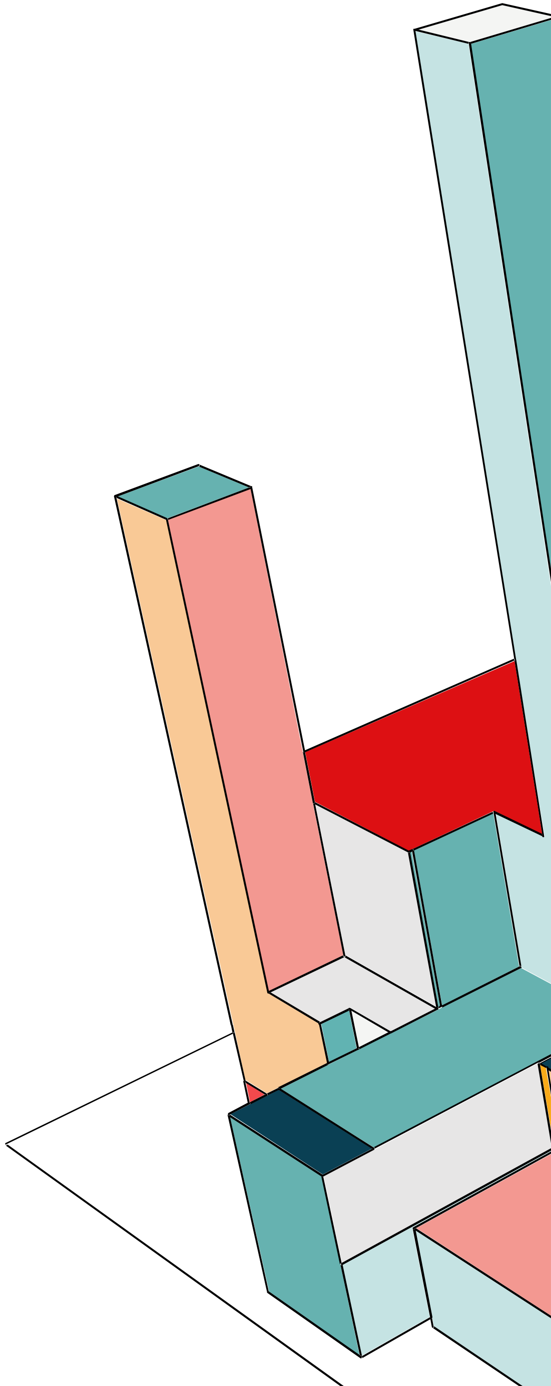
LIGHT MENTORING

Mentors

- Federico D'Asaro Federico.dasaro@linksfoundation.com
- Juan José Márquez Villacís juan.marquez@linksfoundation.com

Weekly one-hour calls with students for the whole duration of the semester

Feel free to reach out via **Slack** or **email** at any time for any questions or doubts



POLICY

- Both project descriptions and implementations will be part of a repository group published on GitHub

- The repositories will be public