Document Validation Report

for OCEANAI

Correspondence: Yes
Percentages: 80.0%

Conclusion:

The repository contains Audio, Video, Text and Multimodal classes whose internal pipelines match the documentation: openSMILE + log-mel + LSTM/VGG for audio, MediaPipe landmarks + ResNet-50/LSTM for video, Whisper + LIWC + BERT + bi-LSTM/Attention for text, and a fusion network that combines the three modalities. Demo notebooks apply the models to the exact datasets named in the paper (First-Impressions V2 and MuPTA) and compute Accuracy/MAE, so inference-time reproduction is possible via the supplied pretrained weights. However, some elements promised in the doc (e.g., explicit YOLO face detection, cross-modal transformer/siamese attention fusion, full training scripts) are not evident; the code focuses on inference only. Because those missing or differently-implemented parts limit full experimental replication, alignment is substantial but not complete.