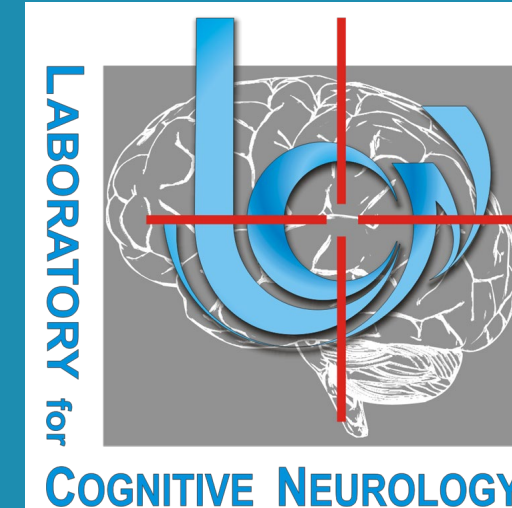


P148. Comparison between auditory, textual, and late-fusion Transformer models for the discrimination of prodromal Alzheimer’s disease based on spontaneous speech



Bastiaan Tamm^{1,2}, Helena Balabin^{1,3}, Kevin Statz¹, Hugo Van hamme², Rik Vandenberghe^{1,4}

1. Laboratory for Cognitive Neurology, Department of Neurosciences, KU Leuven, Belgium; 2. Processing Speech and Images, Department of Electrical Engineering, KU Leuven, Leuven, Belgium; 3. Language Intelligence and Information Retrieval Lab, Department of Computer Science, KU Leuven, Leuven, Belgium; 4. Neurology service, University Hospitals Leuven, Leuven, Belgium

Background

Interest in **digital assessment of spontaneous speech** for clinical trials is rising given its scalability, ecological validity, and the revolutionary progress in Natural Language Processing models. However, it remains unclear whether the auditory signal, the transcript, or a late fusion of both provides the most reliable basis for detecting prodromal Alzheimer’s disease.

Methods

	Amyloid Neg. Control	Prodromal AD
N	63	53
Age (year)	72.3 ± 6.9	71.6 ± 6.3
Sex (f:m)	35:28	23:30
CDR (0:0.5:1)	63:0:0	3:42:8
MMSE	28.8 ± 1.0	25.4 ± 3.1

Spontaneous speech data collection

1. Autobiographical interview (Levine, 2004)
- For five periods of their life, participants were asked to describe one life event of their choice in detail
- Early childhood [5-11 years]

• Adolescent teen age years [12-17 years]

• Early adulthood [18-35 years]

• Middle age [36-55 years]

• Most recent year
2. Describe a day of the past week chronologically from morning to evening
3. Describe a number of recent news events in detail
4. Noun description: 20 abstract and 20 concrete
5. Cookie Theft description

Automated Speech Recognizer

In-house model (Poncelet & Van Hamme, 2023)

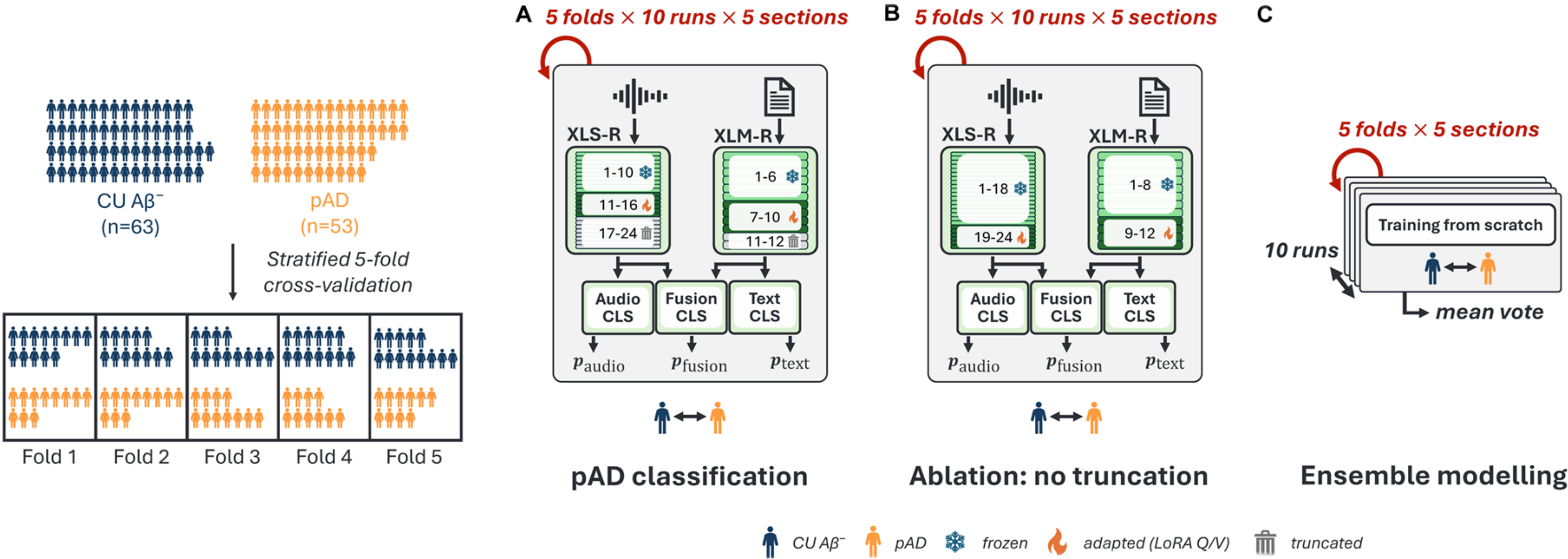
Analysis of diagnostic accuracy

Receiver Operating Characteristics (ROC) analysis with 5-fold cross-validation (CV)

Contact

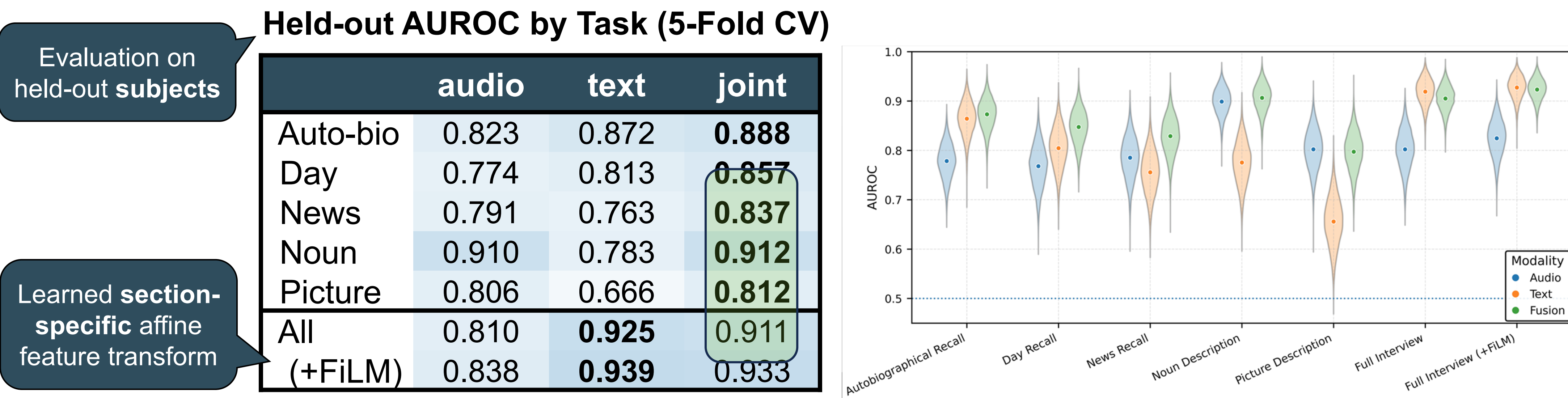
bastiaan.tamm@kuleuven.be
rik.vandenberghe@uzleuven.be

Model and classifier



Results

1. Whether classification based on auditory or textual input performs best, depends on the task performed. Regardless of the task, late fusion always performs at least equally well as the best performing unimodal model.
2. Diagnostic accuracy of the text-based model and the late-fusion model is high.



Interpretation

Tasks that are more constrained, such as *picture description* and *noun description*, are best modeled based on the auditory modality: Linguistic content is less discriminating than how it is spoken under these circumstances. On the other hand more open-ended tasks, such as the *autobiographical interview*, are best modeled based on the textual modality as the lexical and semantic features contain more diagnostic information under such circumstances.

Conclusion

High diagnostic performance can be achieved by late fusion models for discriminating early clinical stages of AD from healthy controls.