

Language-based Dementia Classification Should Consider Model Cognition For Interpretability

Yui Ishihara¹, Michelle Cohn^{2,3,4}, Kartik Patwari¹, Alyssa Weakley⁴, Chen-Nee Chuah¹

¹UC Davis, Electrical and Computer Engineering, ²UC Davis Linguistics,
³UCSF Memory & Aging Center, ⁴UC Davis Alzheimer's Disease Research Center

Background

- Speech and language can be sensitive indicators of cognitive decline. [1]
- Traditional ML and LLMs are increasingly being applied to dementia detection tasks using speech transcripts. [2]

Current Limitations

- Many works collapse progressive cognitive decline into a single classification task. [3, 4]
- Limited transparency regarding what features the model uses.
- Current datasets typically only contain:
 - diagnosis labels, transcripts, audio files [2]

! High accuracy does not necessarily result in clinical trustworthiness, and LLMs often operate as black-box systems.

Directions for Future Research

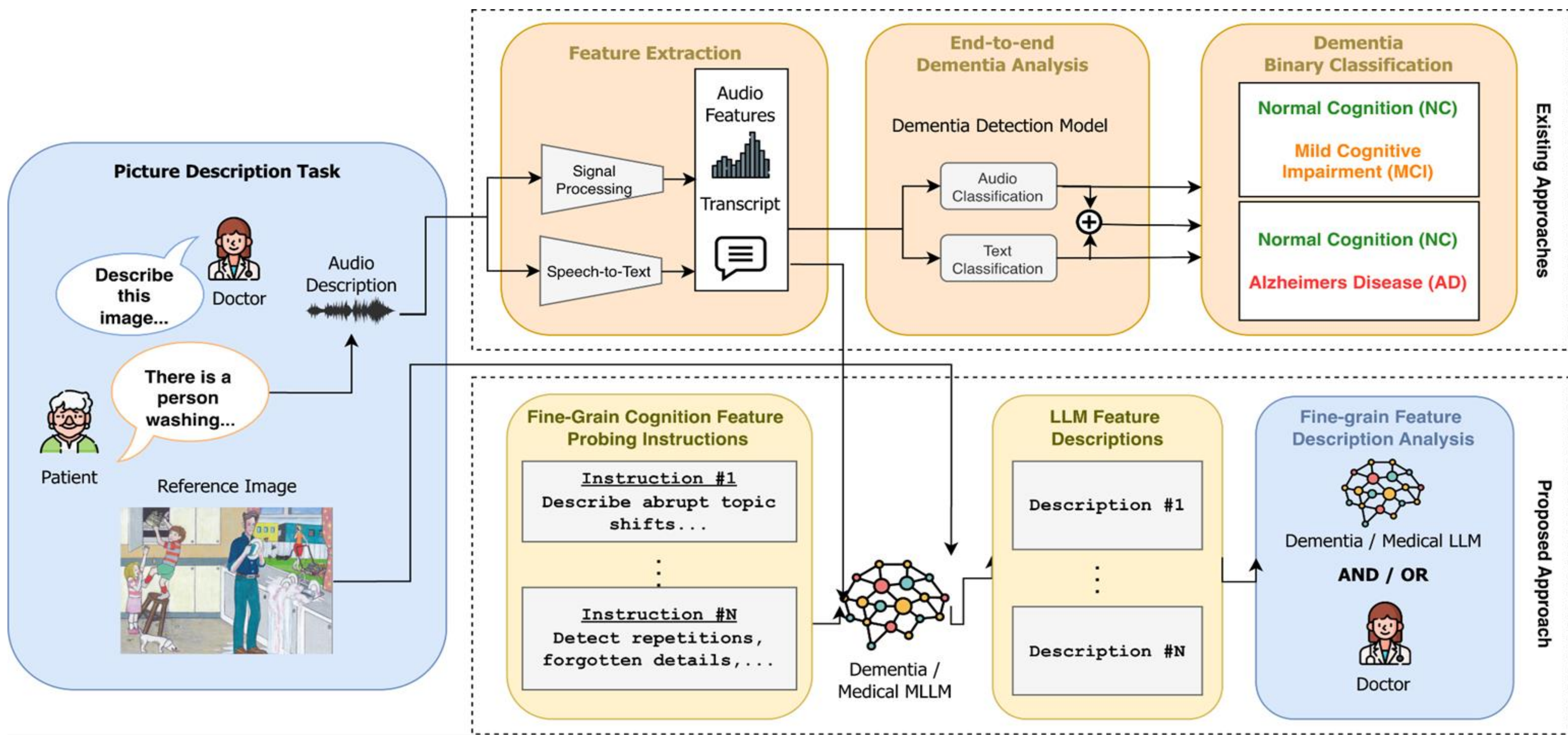
1. Move Beyond Binary Classification

- Modeling **cognitive severity** and **progression**
 - Further study MCI detection [5], MMSE scores.
- Differentiate **disease subtypes** and specific neurodegenerative conditions.
 - e.g., amnesic vs. non-amnesic MCI, Parkinson's disease, Lewy body disease.
- Future systems should **model progression** and granularity to support longitudinal monitoring.

Acknowledgement

This material is based upon work supported by funding from CITRIS and the Banatao Institute at the University of California and a UC Davis Academic Federation Grant.

Conceptual Framework



2. Develop Richer Datasets

- Future datasets should aim to include:**
 - expert-defined linguistic and cognitive **discourse-level annotations**
 - attempt to capture memory, discourse, and reasoning-related features
 - provide **expert supervision** through clinician-derived annotations
 - report detailed **demographic data** to improve fairness, generalizability, and bias evaluation.
 - Expand dataset coverage**
 - languages, tasks, populations, conditions.

3. Probe Model Cognition

- Explore which **features** drive model predictions
 - report interpretability alongside accuracy
- Quantitatively benchmark whether models rely on **clinically meaningful features**
 - e.g. include sensitivity, specificity, AUROC
- Compare model reasoning with **clinician reasoning** for generative outputs

References

- Saturnino Luz, Fasih Haider, ... and Brian MacWhinney. Alzheimer's dementia recognition through spontaneous speech, 2021.
- Zhu, et al. Adversarial text generation using large language models for dementia detection. EMNLP, 2024.
- Uddin, et al. A novel approach utilizing machine learning for the early diagnosis of alzheimer's disease. Biomedical Materials & Devices, 2023.
- Benjamin Barrera-Altuna, ... and Seungbae Kim. Multilingual mild cognitive impairment detection with multimodal approach. 2024.
- Pérez-Toro, et al. Multilingual speech and language analysis for the assessment of mild cognitive impairment: Interspeech 2024.

Transcript	Label	Missing Details?	Repetition?	Disorganized Speech?	Clarity Issues?	Clinician Notes
"...girl uh stealing cookies..."	MCI	✓	✓	X	X	Missing minor details
"...mother washing dishes..."	NC	X	✓	X	X	Complete narrative
"...he's reaching under..."	AD	✓	✓	✓	✓	Incorrect key scene details