EPiC Series in Technology

Volume 3, 2026, Pages 198–207

Proceedings of AI for Accelerated Research Symposium

Real-World Car Talk: A Digital Biomarker of Cognitive Trajectory

Aparna Joshi[1], Matthew Rizzo[2], and Anuj Sharma[1]

[1] Department of Civil, Construction and Environmental Engineering, Iowa State University, Ames, IA, USA

[2] Department of Neurological Sciences, University of Nebraska Medical Center, Omaha, NE, USA

Abstract

Early detection of mild dementia is vital for timely intervention, yet most existing voice biomarker research relies on scripted, high-quality clinical recordings. Such settings fail to capture the acoustic variability of everyday life, limiting real-world applicability. This proof-of-concept study is the first to develop a voice-based biomarker for mild dementia using *naturalistic* in-vehicle audio data collected during routine driving. Audio recordings from 29 participants with sufficient speech content were processed to isolate speech segments through unsupervised clustering, followed by manual verification to ensure relevance. Speech embeddings were extracted using the Wav2Vec2 architecture and a multilayer perceptron classifier was trained and evaluated in a subject-level leave-one-subject-out (LOSO) framework. The model achieved an accuracy of 68.97%, precision of 75%, recall of 60%, and F1-score of 66.67%. These findings demonstrate the feasibility of deriving robust voice biomarkers from highly variable, noise-rich real-world audio. This work lays the groundwork for scalable, passive in-vehicle cognitive health monitoring, with future directions including larger datasets, multimodal integration, and longitudinal analysis for early dementia detection.

1 Introduction

Dementia currently affects more than 50 million people worldwide, and this number is projected to triple by 2050 [1]. Alzheimer's disease (AD) is the most prevalent form of dementia, accounting for an estimated 60-80% of all cases [2]. Mild Cognitive Impairment (MCI) represents the prodromal stage of transition between healthy aging and dementia, characterized by subtle cognitive changes that do not yet meet the criteria for a major neurocognitive disorder [3]. Early identification of individuals in the prodromal phase could enable timely detection and help reduce the burden of AD on both patients and caregivers [4, 5].

Consequently, there is a growing emphasis on developing innovative tools and biomarkers for the early detection of the pre-dementia stages of AD. Traditional biomarker approaches, including neuro-imaging and invasive fluid sampling, while effective, are costly and impractical for broad population screening [6, 7]. In AD, progressive brain pathology disrupts sensory, motor, and cognitive processes, often manifesting as impaired spontaneous speech [8, 9]. Evidence shows that AD patients often exhibit slower speech, longer pauses, and increased word-finding difficulty, leading to disrupted fluency [10]. These findings suggest the potential for developing more accurate AD risk prediction models based on vocal features.

Several studies have explored voice- and audio-based biomarkers for cognitive impairment, using acoustic, linguistic, and prosodic cues to distinguish AD, MCI, and healthy aging [11, 12]. While many of these models achieve strong performance (99% accuracy), they are typically trained on data collected in controlled settings with scripted speech or structured interviews [13]. Such conditions minimize variability and noise but reduce robustness to spontaneous speech and real-world environments, limiting their applicability for passive monitoring.

There is growing recognition of the need for evidence derived from everyday speech in highly variable, real-world environments. Naturalistic Driving Studies (NDS) offer an inherently uncontrolled setting, capturing real-world interactions between drivers, vehicles, and the environment. These studies have been widely used to investigate driver behavior, safety risks, and cognitive decline, leveraging measures such as speed variability, lane-keeping, reaction times, and trip patterns to detect early signs of impairment [14, 15, 16]. Because NDS reflect everyday driving behavior, they provide a unique opportunity to study cognitive function in ecologically valid contexts [17, 18]). However, voice data incidentally captured during these in-vehicle recordings remains largely unexplored as a source of cognitive health biomarkers.

In this proof-of-concept study, we leverage these naturalistic, driving audio recordings to detect mild dementia and mild cognitive impairment in older drivers using a deep learning approach. Our method combines speech activity detection, speaker diarization, and clustering to isolate and identify the driver's speech from noisy, real-world in-vehicle recordings. Speech embeddings were extracted using the Wav2Vec2 architecture and classified with a multilayer perceptron (MLP) to distinguish mild dementia and MCI from healthy aging. The model was evaluated using subject-level leave-one-subject-out cross-validation (LOSO-CV) to minimize bias and ensure generalizability. This approach demonstrates the feasibility of using in-vehicle audio for passive, ecologically valid biomarkers of early cognitive impairment.

2 Dataset Description

This study used in-vehicle audio data extracted from in-cabin video recordings of 40 older adult drivers (65 to 90 years) participating in a larger NDS conducted in Omaha, Nebraska, between March 2021 and June 2024. Participants were legally licensed drivers living independently in Nebraska or nearby areas, able to communicate in English, and meeting state vision requirements (visual acuity $\geq$ 20/40, corrected or uncorrected). Individuals with medical conditions or medications known to significantly affect cognition or alertness were excluded. Written informed consent was obtained under the University of Nebraska Medical Center IRB approval (IRB #522-20-FB).

Data Collection: A custom in-vehicle sensor system ("Black Box") was installed in each participant's personal vehicle to continuously record naturalistic driving behavior over a three-month baseline period. The system activated automatically at engine ignition and deactivated when the vehicle was turned off, with each ignition cycle logged as a separate "drive." Data streams included forward roadway video, in-cabin video, GPS, inertial measurement unit (IMU) data, and on-board diagnostics (OBD) parameters (e.g., vehicle speed, throttle position, engine RPM). For the purposes of this study, only the audio channel from the in-cabin video was used for analysis.

Cognitive Classification: Participants were classified as healthy aging (HA), having mild cognitive impairment, or mild dementia based on the 2018 National Institute on Aging–Alzheimer's Association (NIA-AA) research framework [19, 20]. All diagnoses were determined through comprehensive neurological and neuropsychological evaluations conducted by two dementia-specialist clinicians, using the National Alzheimer's Coordinating Center (NACC) Uniform Data Set (UDS) battery [21]. Functional status was assessed using the Functional Activities Questionnaire (FAQ) (Marshall et al., 2015) and the Clinical Dementia Rating (CDR) scale [22].

This proof-of-concept analysis included 40 older drivers across three cognitive groups: healthy aging (n = 20), mild cognitive impairment (n = 18), and mild dementia (n = 2). Due to the small number of mild dementia cases, participants with mild cognitive impairment and mild dementia were combined into a single category labeled '*cognitive impairment*' for analysis. For each participant, audio was extracted from in-cabin video recordings, which were stored in discrete 1-minute video files, resulting in approximately 30 files per participant. These segments captured all in-cabin activity during that period and could contain varying amounts of driver speech, background conversation, or silence. The presence and duration of speech in each segment were determined later during the speech activity detection stage of our processing pipeline.

3 Methods

3.1 Data Processing Pipeline

The audio segments were processed through a multi-stage pipeline to isolate driver speech and generate model-ready embeddings.

Speech Activity Detection (SAD): Speech activity detection was first applied using the Pyannote.audio pre-trained SAD model [23] to identify regions containing speech and remove silence or background noise. This filtering step reduced computational load and ensured that downstream processing was applied only to segments with detected speech.

Speaker Diarization: The speech-containing time intervals identified in the SAD stage were then processed with Pyannote.audio's diarization pipeline to assign speaker labels to each speech activity segment. In the naturalistic driving environment, multiple audio sources were often present, including passengers, the driver, and background elements such as the radio or navigation prompts. The diarization model assigned numerical speaker IDs within each 1-minute audio file independently, meaning that the same person could receive different speaker IDs across 1-minute audio files. A key challenge, therefore, was determining which of these diarized speakers corresponded to the driver across all recordings for a given participant.

Driver Identification and Verification: To identify the driver, we first extracted speaker embeddings from all diarized speech segments lasting at least 0.5 seconds. For each participant, embeddings were clustered to group similar voices across multiple 1-minute files, resulting in an average of 4–5 clusters per participant, depending on the presence of background noise and other co-passengers. We then manually reviewed the clusters for each participant to identify the main driver cluster, ensuring accurate identification before downstream analysis. This driver identification process was performed for all 40 participants, with manual verification serving as a quality control step. Following this step, we found that 6 participants in the healthy aging group and 5 participants in the MCI group had no speech across their recordings. Therefore, the final dataset used for model training and evaluation included 29 participants (MCI/mild dementia = 15; healthy aging=14) with verified driver speech. Across these participants, the total speaking time (after stitching all verified speech segments) ranged from approximately 20 seconds to 20 minutes per subject. The age distribution for the healthy aging group had a mean of 75.5 years (SD = 5.72, range = 68–88), while the MCI/Mild dementia group had a mean age of 77.8 years

(SD = 8.37, range = 65–90). Sex distributions were unequal across groups, with the healthy aging group comprising 71% female and 29% male participants, whereas the MCI/MD group comprised 73% male and 27% female participants. Given the observed age and sex imbalances across groups, we used a self-supervised speech model (Wav2Vec2) to extract acoustic-phonetic features. This approach reduces the risk of model predictions being confounded by demographic characteristics, as it focuses on speech content rather than speaker identity.

Feature Extraction - Wav2Vec2 Embeddings: Verified driver speech segments were passed through the pre-trained Wav2Vec2 model [2] using the HuggingFace Transformers library. Wav2Vec2 is a self-supervised model trained on unlabelled speech (LibriSpeech) to learn rich acoustic and phonetic representations directly from raw audio. Unlike traditional spectral features such as Mel-Frequency Cepstral Coefficients (MFCCs), Wav2Vec2 embeddings capture contextualized speech characteristics without requiring explicit transcription or word-level labels, making them robust to variations in content, speaking style, and background noise. Audio recordings were upsampled to 16 kHz to meet model input requirements, and the model produced 512-dimensional embeddings representing high-level acoustic context. These embeddings served as the sole input for downstream voice-based classification models.

3.2 Model Architecture and Evaluation

Classifier Architecture: Speech segments identified as belonging to the driver were first divided into fixed-length 5-second chunks. For each chunk, 192-dimensional speaker embeddings were extracted using the Wav2Vec2 model. Embeddings were L2-normalized prior to classification. These embeddings served as input to a multilayer perceptron classifier implemented in PyTorch. An MLP was chosen for downstream classification due to its ability to model non-linear decision boundaries in embedding space without requiring large amounts of data for training. More complex sequence models (e.g., LSTMs or Transformers) were avoided in this proof-of-concept phase to minimize overfitting risk, given the modest sample size (n = 29). The MLP architecture strikes a balance between expressiveness and computational efficiency, making it well-suited for potential edge-device deployment in real-time in-vehicle monitoring. The classifier architecture consisted of:

Input layer: Fully connected linear layer mapping the embedding dimension (Wav2Vec2 output) to 256 hidden units, followed by a ReLU activation and dropout (p = 0.3).

Hidden layer: Fully connected layer mapping 256 units to 128 units, followed by a ReLU activation and dropout (p = 0.3).

Output layer: Fully connected layer mapping 128 units to n classes (in our case, 2 classes: healthy aging vs. cognitive impairment).

The network was trained with the Adam optimizer (learning rate = 1×10^{-4}, weight decay = 1×10^{-4}) and cross-entropy loss for 30 epochs. Mini-batches of size 128 were used.

Training and Evaluation Strategy: We adopted a Leave-One-Subject-Out Cross-Validation (LOSO-CV) framework, where in each fold, one participant's data was held out for testing while the remaining participants' data was used for training. To balance training data, 70 chunks (each chunk 5 seconds) were randomly sampled per subject with replacement. For testing, all available chunks from the held-out participant were used. During testing, all available audio chunks were used.

Performance Metrics: At test time, the classifier generated predictions for each chunk, and subject-level predictions were obtained via ensemble voting across all chunk predictions. Classification performance was reported at the subject level in terms of accuracy, precision, recall, and F1-score. Additionally, we reported the area under the curve (AUC) and the confusion matrix for subject-level predictions. Figure 1 illustrates the proposed pipeline for detecting cognitive impairment from in-vehicle speech, from audio capture to final subject-level classification.

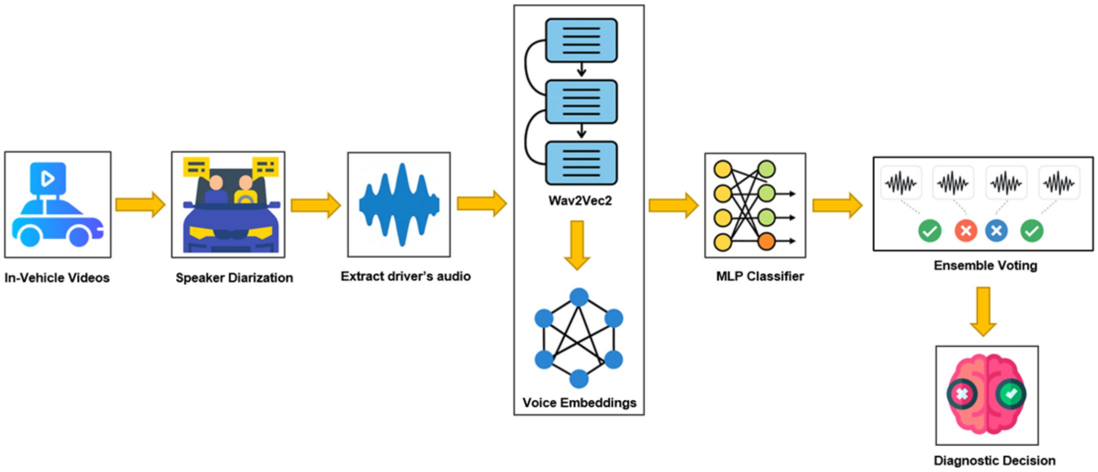


Figure 1: Overview of the proposed framework for cognitive impairment detection using in-vehicle speech. Raw driving audio is processed through speech activity detection and diarization, followed by embedding extraction using a Wav2Vec2 model. The embeddings are classified with a multilayer perceptron, and predictions from multiple audio chunks are aggregated via ensemble voting to generate a subject-level decision.

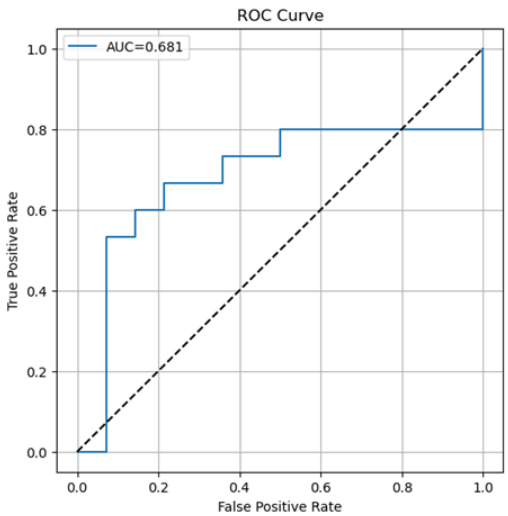
4 Results

The proposed Wav2Vec2 + MLP classifier achieved an accuracy of 68.97%, precision of 75%, recall of 60%, and F1-score of 66.67% at the subject level. In this context, "subject level" means that instead of evaluating the model on individual 5-second audio chunks, we combined all chunk-level predictions from each participant to produce one classification per person. Out of 29 participants, 20 were correctly classified. The area under the Receiver Operating Characteristic curve (AUCROC) was 0.68, indicating good discrimination between the two classes as shown in Figure 2a. The step-like pattern in the ROC curve is expected, given the discrete nature of subject-level predictions and the limited sample size (n = 29), where each point represents an entire participant. The confusion matrix (Figure 2b) illustrates that correct classification rates were similar across both groups, with only minor misclassification between healthy aging and cognitive impairment. Specifically, 11 out of 14 cognitively healthy participants and 11 out of 15 participants with cognitive impairment were classified correctly.

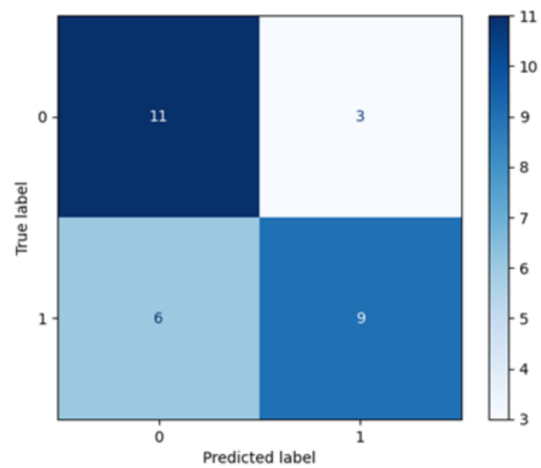
The training loss curve in Figure 2c demonstrates consistent model convergence over 30 epochs, with no signs of overfitting, as indicated by the smooth downward trajectory and the plateau near zero loss.

5 Discussions

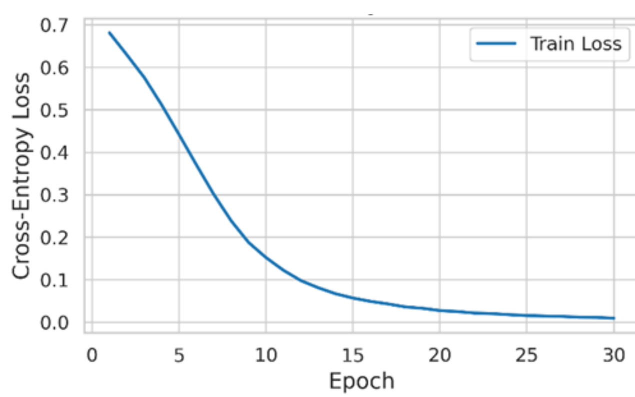
This study demonstrates the feasibility of detecting cognitive impairment using voice features extracted from naturalistic, in-vehicle audio recordings. Unlike most prior voice biomarker



(a) Receiver Operating Characteristic curve



(b) Confusion Matrix



(c) Training loss curve averaged across folds

Figure 2: Model performance of Wav2Vec2 + MLP under LOSO-CV.

research, which relies on scripted speech in controlled clinical settings or on structured telephone conversations, our dataset was collected in real driving environments, with significant variability in background noise, microphone placement, and speech content. Despite these challenges, the Wav2Vec2 + MLP framework achieved a subject-level accuracy of 68.97%, precision of 75%, recall of 60%, F1 of 66.67%, and an AUC of 0.68 under a leave-one-subject-out (LOSO) evaluation. These results suggest that speaker embedding-based models can retain discriminative information even under adverse acoustic conditions.

A notable strength of this approach is its applicability to passive cognitive health monitoring scenarios. In-vehicle speech is a naturally occurring signal that does not require participant effort or specialized equipment beyond what may already exist in modern vehicles. The embedding-based pipeline is lightweight enough to be implemented on edge devices, enabling real-time analysis without compromising privacy by transmitting raw audio.

However, several factors influenced model performance. Misclassifications were often associated with participants who contributed very short speech durations, segments with high levels of overlapping talkers, or speech dominated by background media such as radio. These issues highlight the importance of robust speech activity detection and diarization in pre-processing.

Additionally, the step-like ROC curve observed in the results reflects the discrete nature of LOSO subject-level predictions and the relatively small sample size. While the present work is a critical proof-of-concept, several limitations remain. The dataset size was modest, and 11 of the 40 recruited participants had no detectable speech, resulting in only 29 participants for analysis. Increasing the subject pool in number will be an essential next step to evaluate generalization across different demographics, vehicle types, and driving conditions. Moreover, the current driver identification process relies on manual verification of diarization clusters to ensure correct assignment, which is feasible for small datasets but impractical at scale. Automating driver identification using advanced clustering validation or supervised speaker identification models will be critical for scaling up the approach.

Compared to prior clinical speech biomarker studies, our accuracy is lower than that reported in controlled settings (often 90- 99%), but the trade-off is increased ecological validity. The real-world noise and variability captured here more closely reflect deployment conditions, suggesting that models trained on such data may generalize better to everyday environments. This proof-of-concept, therefore, offers an important bridge between laboratory research and real-world application.

Future work will expand the dataset to a larger, demographically balanced cohort of 118 participants, enabling more robust statistical evaluation and the training of end-to-end models fine-tuned for in-vehicle speech. Additionally, we plan to extract voice biomarkers from spoken neuro-psychological assessments conducted in the clinic. This will allow us to evaluate the consistency and ecological validity of in-vehicle speech relative to structured clinical recordings, and to determine whether naturalistic driving speech better reflects cognitive status.We will also incorporate longitudinal clinical voice recordings to track temporal changes in speech patterns that may precede clinically detectable cognitive decline. Integrating additional modalities, such as driving performance metrics or physiological data, can further improve the accuracy of the detection.

6 Author Contributions

A. Joshi developed the methodology, conducted the formal analysis, implemented the computational pipeline, and prepared the original draft of the manuscript. M. Rizzo provided access to the data and contributed to editing and reviewing the manuscript. A. Sharma served as the project advisor, reviewed the results, and contributed to editing and reviewing the manuscript. All authors read and approved the final manuscript.

7 Acknowledgments

We extend our sincere gratitude to the Mind and Brain Health Labs in the Department of Neurological Sciences at UNMC for their leadership in recruiting participants for this study.

8 Data Statement

Given the sensitive nature of the data collected in this study, participants were assured that the raw data would remain confidential and would not be disclosed.

References

[1] World Health Organization. Global action plan on the public health response to dementia, 2017.

[2] Alexei Baevski, Henry Zhou, Abdelrahman Mohamed, and Michael Auli. wav2vec 2.0: A framework for self-supervised learning of speech representations. *arXiv*, 2020. https://arxiv.org/abs/2006.11477.

[3] Cai Gillis, Fariba Mirzaei, Michele Potashman, M. Arfan Ikram, and Nancy Maserejian. The incidence of mild cognitive impairment: A systematic review and data synthesis. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 11(1):248–256, 2019. https://doi.org/10.1016/j.dadm.2019.01.004.

[4] Andrea Bradford, Christina Upchurch, David Bass, Katherine Judge, A. Lynn Snow, Nancy Wilson, and Mark E. Kunik. Knowledge of documented dementia diagnosis and treatment in veterans and their caregivers. *American Journal of Alzheimer's Disease & Other Dementias*, 26(2):127–133, 2011. https://doi.org/10.1177/1533317510394648.

[5] Bruno Dubois, Alessandro Padovani, Philip Scheltens, Andrea Rossi, and Grazia Dell'Agnello. Timely diagnosis for alzheimer's disease: A literature review on benefits and challenges. *Journal of Alzheimer's Disease*, 49(3):617–631, 2015. https://doi.org/10.3233/JAD-150692.

[6] Spencer A. W. Lee, Luciano A. Sposato, Vladimir Hachinski, and Lauren E. Cipriano. Cost-effectiveness of cerebrospinal biomarkers for the diagnosis of alzheimer's disease. *Alzheimer's Research & Therapy*, 9(1):18, 2017. https://doi.org/10.1186/s13195-017-0243-0.

[7] Frank S. Prato, William F. Pavlosky, Steven C. Foster, Jonathan D. Thiessen, and Roderic P. Beaujot. Screening for dementia caused by modifiable lifestyle choices using hybrid pet/mri. *Journal of Alzheimer's Disease Reports*, 3(1):31–45, 2019. https://doi.org/10.3233/ADR-180098.

[8] Clive Ballard, Serge Gauthier, Anne Corbett, Carol Brayne, Dag Aarsland, and Emma Jones. Alzheimer's disease. *The Lancet*, 377(9770):1019–1031, 2011. https://doi.org/10.1016/S0140-6736(10)61349-9.

[9] Kamalini G. Ranasinghe, Jeevit S. Gill, Hardik Kothare, Alexander J. Beagle, Danielle Mizuiri, Susanne M. Honma, Maria Luisa Gorno-Tempini, Bruce L. Miller, Keith A. Vossel, Srikantan S. Nagarajan, and John F. Houde. Abnormal vocal behavior predicts executive and memory deficits in alzheimer's disease. *Neurobiology of Aging*, 52:71–80, 2017. https://doi.org/10.1016/j.neurobiolaging.2016.12.020.

[10] F. Martínez-Sánchez, J. J. G. Meilán, J. García-Sevilla, J. Carro, and J. M. Arana. Análisis de la fluencia lectora en pacientes con la enfermedad de alzheimer y controles asintomáticos. *Neurología*, 28(6):325–331, 2013. https://doi.org/10.1016/j.nrl.2012.07.012.

[11] Ihab Hajjar, Maureen Okafor, Jinho D. Choi, Elliot Moore, Anees Abrol, Vince D. Calhoun, and Felicia C. Goldstein. Development of digital voice biomarkers and associations with cognition, cerebrospinal biomarkers, and neural representation in early alzheimer's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 15(1), 2023. https://doi.org/10.1002/dad2.12393.

[12] Akihiro Shimoda, Yue Li, Hana Hayashi, and Naoki Kondo. Dementia risks identified by vocal features via telephone conversations: A novel machine learning prediction model. *PLOS ONE*, 16(7):e0253988, 2021. https://doi.org/10.1371/journal.pone.0253988.

[13] Máté Hireš, Matej Gazda, Peter Drotár, Nemuel Daniel Pah, Mohammod Abdul Motin, and Dinesh Kant Kumar. Convolutional neural network ensemble for parkinson's disease detection from voice recordings. *Computers in Biology and Medicine*, 141:105021, 2022. https://doi.org/10.1016/j.compbiomed.2021.105021.

[14] Aparna Joshi, Kojo Adugyamfi, Jennifer Merickel, Pujitha Gunaratne, and Anuj Sharma. Driver age and its effect on key driving metrics: Insights from dynamic vehicle data. *arXiv*, 2025. https://doi.org/10.48550/arXiv.2501.06918.

[15] Aparna Joshi, Jennifer Merickel, Cyrus V. Desouza, Matthew Rizzo, Pujitha Gunaratne, and Anuj Sharma. Investigating speed deviation patterns during glucose episodes: A quantile regression approach. In *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*, pages 5750–5755, 2023. https://doi.org/10.1109/ITSC57777.2023.10421822.

[16] Aparna Joshi, Archana Venkatachalapathy, Jennifer Merickel, Jun Ha Chang, Matthew Rizzo, Soumik Sarkar, and Anuj Sharma. Impact of acute glucose episodes on adherence to speed limits in the naturalistic setting for drivers with diabetes: An application of linear quantile mixed models. *Transportation Research Interdisciplinary Perspectives*, 34:101633, 2025. https://doi.org/10.1016/j.trip.2025.101633.

[17] Aparna Joshi, Guillermo Basulto-Elias, Souradeep Chattopadhyay, Anuj Sharma, Soumik Sarkar, Daniel Murman, Jun Ha Chang, and Matthew Rizzo. AI on the road: Detecting dementia with real-world behavior data. In *150th Annual Meeting American Neurological Association, Annals of Neurology*, volume 98, pages S1–S358, 2025. https://doi.org/10.1002/ana.78015.

[18] Aparna Joshi, Jun Ha Chang, Guillermo Basulto-Elias, Shauna Hallmark, Matthew Rizzo, and Anuj Sharma. Identifying and predicting cognitive decline using multi-modal sensor data and machine learning approach. *Research Square*, 2025. https://doi.org/10.21203/rs.3.rs-6735622/v1.

[19] Clifford R. Jack, David A. Bennett, Kaj Blennow, Maria C. Carrillo, Billy Dunn, Samantha Budd Haeberlein, David M. Holtzman, William Jagust, Frank Jessen, Jason Karlawish, Enchi Liu, Jose Luis Molinuevo, Thomas Montine, Creighton Phelps, Katherine P. Rankin, Christopher C. Rowe, Philip Scheltens, Eric Siemers, Heather M. Snyder, and Nina Silverberg. Nia–aa research framework: Toward a biological definition of alzheimer's disease. *Alzheimer's & Dementia*, 14(4):535–562, 2018. https://doi.org/10.1016/j.jalz.2018.02.018.

[20] Clifford R. Jack, J. Scott Andrews, Thomas G. Beach, Teresa Buracchio, Billy Dunn, Ana Graf, Oskar Hansson, Carole Ho, William Jagust, Eric McDade, Jose Luis Molinuevo, Ozioma C. Okonkwo, Luca Pani, Michael S. Rafii, Philip Scheltens, Eric Siemers, Heather M. Snyder, Reisa Sperling, Charlotte E. Teunissen, and Maria C. Carrillo. Revised criteria for diagnosis and staging of alzheimer's disease: Alzheimer's association workgroup. *Alzheimer's & Dementia*, 20(8):5143–5169, 2024. https://doi.org/10.1002/alz.13859.

[21] Marilyn S. Albert, Steven T. DeKosky, Dennis Dickson, Bruno Dubois, Howard H. Feldman, Nick C. Fox, Anthony Gamst, David M. Holtzman, William J. Jagust, Ronald C.

Petersen, Peter J. Snyder, Maria C. Carrillo, Bill Thies, and Creighton H. Phelps. The diagnosis of mild cognitive impairment due to alzheimer's disease: Recommendations from the national institute on aging–alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. *Alzheimer's & Dementia*, 7(3):270–279, 2011. https://doi.org/10.1016/j.jalz.2011.03.008.

[22] John C. Morris. The clinical dementia rating (cdr). *Neurology*, 43(11):2412, 1993. https://doi.org/10.1212/WNL.43.11.2412-a.

[23] Hervé Bredin. pyannote.audio 2.1 speaker diarization pipeline: principle, benchmark, and recipe. In *Proc. INTERSPEECH 2023*, 2023.