Detection of mild cognitive impairment from non-semantic, acoustic voice features: the Framingham Heart Study

Huitong Ding, Adrian Lister, Cody Karjadi, Rhoda Au, Honghuang Lin, Brian Bischoff, Phillip H. Hwang

Submitted to: JMIR Aging
on: December 03, 2023

Table of Contents

Detection of mild cognitive impairment from non-semantic, acoustic voice features: the Framingham Heart Study

Huitong Ding[1] PhD; Adrian Lister[2]; Cody Karjadi[1]; Rhoda Au[1]; Honghuang Lin[3]; Brian Bischoff[4]; Phillip H. Hwang[5]

[1]Boston University Chobanian & Avedisian School of Medicine Boston US
[2]Headwaters Innovation, Inc Inver Grove Heights US
[3]University of Massachusetts Chan Medical School Worcester US
[4]Headwaters Innovation, Inc. Inver Grove Heights US
[5]Boston University School of Public Health Boston US

Corresponding Author:
Huitong Ding PhD
Boston University Chobanian & Avedisian School of Medicine
72 E Concord St
Boston
US

Abstract

Background: With the aging global population and the rising burden of Alzheimer's disease and related dementias (ADRD), there is a growing focus on identifying mild cognitive impairment (MCI) to enable timely interventions that could potentially slow down the onset of clinical dementia. The production of speech by an individual is a cognitively complex task that engages various cognitive domains. The ease of audio data collection highlights the potential cost-effectiveness and noninvasive nature of using human speech as a tool for cognitive assessment.

Objective: This study aimed to construct a machine learning pipeline that incorporates speaker diarization, feature extraction, feature selection, and classification, to identify a set of acoustic features derived from voice recordings that exhibit strong MCI detection capability.

Methods: The study included 100 MCI cases and 100 cognitively normal (CN) controls matched for age, sex, and education from the Framingham Heart Study. Participants' spoken responses to neuropsychological test questions were recorded, and the recorded audio was processed to identify segments of each participant's voice from recordings that included voices of both testers and participants. A comprehensive set of 6385 acoustic features was then extracted from these voice segments using the OpenSMILE and Praat softwares. Subsequently, a random forest model was constructed to classify cognitive status using the features that exhibited significant differences between the MCI and CN groups. The MCI detection performance of various audio lengths was further examined.

Results: An optimal subset of 29 features were identified that resulted in an area under the receiver operating characteristic curve (AUC) of 0.87, with a 90% confidence interval from 0.82 to 0.93. The most important acoustic feature for MCI classification was the number of filled pauses (importance score = 0.09). There was no substantial difference in performance of the model trained on the acoustic features derived from different lengths of voice recordings.

Conclusions: This study showcases the potential of monitoring changes to non-semantic and acoustic features of speech as a way of early ADRD detection and motivates future opportunities for using human speech as a measure of brain health.

(JMIR Preprints 03/12/2023:55126)
DOI: https://doi.org/10.2196/preprints.55126

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✔ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Detection of mild cognitive impairment from non-semantic, acoustic voice features: the Framingham Heart Study

Huitong Ding[1,2], Adrian Lister[3], Cody Karjadi[1,2], Rhoda Au[1,2,4,5], Honghuang Lin[6], Brian Bischoff[3], Phillip H. Hwang[2,4,#]

[1]Department of Anatomy and Neurobiology, Boston University Chobanian & Avedisian School of Medicine, Boston, MA, USA; [2]The Framingham Heart Study, Boston University Chobanian & Avedisian School of Medicine, Boston, MA, USA; [3]Headwaters Innovation, Inc., Inver Grove Heights, MN , USA; [4]Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA; [5]Slone Epidemiology Center and Departments of Neurology and Medicine, Boston University Chobanian & Avedisian School of Medicine, Boston, MA, USA; [6]Department of Medicine, University of Massachusetts Chan Medical School, Worcester, MA, USA.

Address for Correspondence:

Phillip H. Hwang, PhD, MPH; phhwang@bu.edu; (617) 358-4049

715 Albany Street, T3E, Boston, MA 02118.

Abstract

Background: With the aging global population and the rising burden of Alzheimer's disease and related dementias (ADRD), there is a growing focus on identifying mild cognitive impairment (MCI) to enable timely interventions that could potentially slow down the onset of clinical dementia. The production of speech by an individual is a cognitively complex task that engages various cognitive domains. The ease of audio data collection highlights the potential cost-effectiveness and noninvasive nature of using human speech as a tool for cognitive assessment.

Objectives: This study aimed to construct a machine learning pipeline that incorporates speaker diarization, feature extraction, feature selection, and classification, to identify a set of acoustic features derived from voice recordings that exhibit strong MCI detection capability.

Methods: The study included 100 MCI cases and 100 cognitively normal (CN) controls matched for age, sex, and education from the Framingham Heart Study. Participants' spoken responses on neuropsychological tests were recorded, and the recorded audio was processed to identify segments of each participant's voice from recordings that included voices of both testers and participants. A comprehensive set of 6385 acoustic features was then extracted from these voice segments using OpenSMILE and Praat software. Subsequently, a random forest model was constructed to classify cognitive status using the features that exhibited significant differences between the MCI and CN groups. The MCI detection performance of various audio lengths was further examined.

Results: An optimal subset of 29 features were identified that resulted in an area under the receiver operating characteristic curve (AUC) of 0.87, with a 95% confidence interval from 0.81 to 0.94. The most important acoustic feature for MCI classification was the number of filled pauses (importance score = 0.09, P = 3.10E-08). There was no substantial difference in

performance of the model trained on the acoustic features derived from different lengths of voice recordings.

Conclusions: This study showcases the potential of monitoring changes to non-semantic and acoustic features of speech as a way of early ADRD detection and motivates future opportunities for using human speech as a measure of brain health.

Keywords: early detection, Alzheimer's disease and related dementias, mild cognitive impairment, digital voice, machine learning

Introduction

Alzheimer's disease and related dementias (ADRD) constitute a significant public health issue, impacting an estimated 6.2 million individuals in the United States, with projections indicating the number of cases to grow to 12.7 million, and 150 million globally, by 2050[1, 2]. Emerging evidence suggests that the functional, psychological, pathological, and physiological alterations associated with ADRD may manifest many years prior to the clinical onset of cognitive dysfunction[3-6]. This increasing awareness has sparked interest in early detection and monitoring of ADRD, with the goal of implementing timely preventive and therapeutic strategies to slow the progression of the disease. Effective as they are in identifying individuals at high risk of ADRD, conventional diagnostic methods, such as cerebrospinal fluid (CSF) biomarkers and neuroimaging, face accessibility limitations primarily due to their high costs[7] and high subject burden. This limits their applicability to other groups, particularly populations in lower-resourced settings, in effectively monitoring the dynamic progression of the disease. Therefore, there is an urgent need for an effective detection method that has much broader and inclusive reach for the early detection of ADRD.

Producing speech is a cognitively complex task that engages various cognitive

domains[8], and the ease of audio data collection underscores the potential cost-effectiveness and noninvasiveness that using human speech-based features may offer to facilitate early identification of cognitive impairment, including mild cognitive impairment (MCI). Studies have indicated that language deficits may manifest in the prodromal stages of cognitive impairment, often years before clinical diagnosis of dementia[9, 10]. Speech, however, is far richer in characterizing cognition that just language. Audio recordings can yield a variety of attributes, encompassing both acoustic and linguistic features. Acoustic features, given their language-independence, have the potential for a broader global applicability. Previous studies from the Framingham Heart Study (FHS) demonstrated significant associations between acoustic features extracted from voice recordings and two primary clinical indices of neurodegeneration: neuropsychological (NP) test performance[11] and brain volumes[12]. Moreover, acoustic-based models can be readily deployed on devices like hand-held recorders, smartphones, tablets and other internet-connected mobile devices, enabling widespread utilization. These characteristics enable voice as a potential digital biomarker for early cognitive impairment monitoring and detection of MCI.

While the use of speech recordings as a novel measure of cognition is still in the early stages of validation, most of the previous studies have relied on a limited set of acoustic features[13-16], potentially constraining the enhancement of early detection capabilities for ADRD. For instance, some studies have concentrated on mel-frequency cepstral coefficients [13, 15], while others have explored a narrow range of temporal and spectral features (such as duration of utterance, number and length of pauses, and F0)[14, 16]. There has been a notable absence of exploration into diverse categories of features including energy, spectral, cepstral, and voicing-related features. Although deep learning has been utilized to investigate these features, its complexity often compromises interpretability. Therefore, there is a need for research to employ more interpretable methods in exploring a richer set of acoustic features

for the detection of MCI. Furthermore, the question of whether extensive voice recordings is necessary to achieve better cognitive assessment performance has not been thoroughly investigated. These issues have significant implications for the widespread, real-world application of speech as a digital data modality for cognitive assessment.

Therefore, the aims of this study were to explore the utility of acoustic features derived from human speech for the identification of MCI and to assess the impact of duration of voice recordings on predictive performance of MCI identification.

Methods

Study population

Initiated in 1948, FHS is a community-based, longitudinal cohort study. This study initially included 605 FHS participants with at least one audio recording who were aged 60 years or older at the time of the NP exam visit where the recordings were collected. Then, a case-control dataset was created consisting of 100 MCI cases and 100 cognitively normal (CN) controls, and matched on age, sex, and education to control for potential confounders and ensure the reliability of the study results. MCI cases were identified through a clinical review conducted by a panel including at least one neurologist and one neuropsychologist based on criteria from the Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV) and the NINCDS-ADRDA[17]. The details of the cognitive status determination can be found in previous studies[15]. The participants were stratified into six age groups, with each group spanning a five-year interval from 60 to 89 years (e.g., 60-64 years, 65-69 years, 70-74 years, 75-79 years, 80-84 years, 85-89 years). Additionally, there was a separate category for individuals aged 90 and above. Study participants were also stratified into four education groups: high school non-graduates, high school graduates, individuals with some college education, and college graduates. Subsequently, controls were selected from the dataset who matched the cases based on age, sex, and education. The earliest collected voice recording from each participant was

included in this analysis.

Ethics Approval

The procedures and protocols of the Framingham Heart Study were approved by the Institutional Review Board of the Boston University Medical Campus, and written informed consent was obtained from all participants.

Voice recordings

FHS has been monitoring cognitive status since 1976, which includes comprehensive NP testing[18]. Since 2005, FHS has digitally recorded all responses to NP test questions that required a voice response, which encompasses the spoken interactions between the tester and the participant. These recordings have been stored in the .wav format and down-sampled to 16 kHz. The current study included digital voice recordings between September 2005 and March 2020.

Machine learning pipeline

This study developed a machine learning pipeline that incorporated speaker diarization, feature extraction, feature selection, and classification to identify a set of acoustic features that exhibited strong MCI detection capability (**Figure 1**).

Speaker Diarization

To accurately analyze the speech of the participants, it is crucial to distinguish between the participant and the tester, and to determine "who spoke when" [19]. This process is known as speaker diarization, which involves segmenting the voice recordings based on the speaker's identity. In this study, the open-source speaker diarization package, pyannote, was utilized to automatically segment each recording into hypothesized utterances from the tester and the participant[20, 21]. Since the NP administration testing process in FHS is standardized, the segmented dominant speaker, based on the duration of the voice recording, was labeled as the participant's speech in this study. These participant segments were combined for subsequent

analysis.

Feature extraction

To extract relevant information from the voice recordings, OpenSMILE software (version 2.1.3) [22] and Praat software[23] were utilized, which facilitated the extraction of a comprehensive set of 6376 features[24] and nine features, respectively. The OpenSMILE feature set used in this study consisted of 65 low-level descriptors (LLDs). These descriptors included energy, spectral, cepstral, and voicing-related features. Each recording was divided into segments of 20 milliseconds using a sliding window approach with a shifting size of 10 milliseconds[25, 26]. The LLDs were extracted from each segment. By allowing for overlaps between successive windows, we were able to facilitate the conservation of information continuity and enable a more precise capture of the signal's dynamics[25, 26]. First-order delta regression coefficients were calculated for all LLDs. A comprehensive set of functionals, such as mean, maximum, minimum, standard deviation (SD) of segment length, and linear regression slope, were applied to extract statistical characteristics from the LLDs and deltas over the full recordings[27-29]. This process provided a concise representation of the acoustic features across the entire recording. As a result of this summarization process, each recording was_represented by a set of 6376 features from OpenSMILE, capturing essential information about the acoustic properties of the audio data. The details of the feature generation process can be found in a prior study[30]. The Praat script was used to generate nine features on syllable nuclei and filled pauses of the voice recordings[31].

Feature selection

First, z-scores were computed for each feature and those with an absolute z-score greater than two were removed as they were considered as outliers. Then, t-tests were used to determine whether there was a significant difference in each feature between the MCI and CN groups. Features that exhibited a significant difference below a P-value threshold of 0.0025 were then

selected to be included in the model.

Classification model

A random forest model was built using a final set of 29 selected features and the performance of the model was evaluated using 10-fold cross-validation. To evaluate the MCI detection performance of the model, the area under the receiver operating characteristic curve (AUC), along with the 95% confidence interval (CI), for the random forest algorithm was obtained. The importance of each feature was computed using an impurity-based approach[32].

Comparison of performance across different audio recording lengths

To investigate the impact of the length of the audio recordings on the MCI classification performance, the first five minutes, 10 minutes, 15 minutes, and 30 minutes of the whole recording for each participant were extracted. Subsequently, the same processing steps were applied to each extracted audio segment, including speaker diarization, feature extraction, and the construction of the MCI classification model.

Results

Cohort descriptive

The study sample included 200 participants, of whom 100 were diagnosed with MCI and the other 100 participants were classified as CN. In the overall sample, the average age was 74 years (SD = 6 years) and 46% (92/200) were female, with the sex distribution (females versus males) equal in both MCI and CN groups. Education in the overall sample was distributed as follows: 18 participants (18/200, 9%) did not graduate from high school, 54 participants (54/200, 27%) were high school graduates, 66 participants (66/200, 33%) had completed some college, and 62 participants (62/200, 31%) held at least a college degree.

Feature selection and detection performance

Table 1 presents the 29 acoustic features significantly associated with cognitive status, as indicated by t-test *P* values less than .0025. The table also displays the importance scores of

these features for the classification of MCI, with higher values indicating greater importance. The most important acoustic feature for MCI classification was the number of filled pauses (nrFP), with an importance score of 0.09. The optimal model was achieved when including these 29 acoustic features that were based on using a z-score cutoff of two and a p-value threshold of .0025 (AUC: 0.87, 95% CI: 0.81-0.94) (**Figure 2**).

Table 1. The optimal acoustic feature set for MCI detection.

Feature	Description	Importance[a]	*P* value[b]
nrFP	Number of filled pauses	0.09	3.10×10^{-8}
tFP	Total time of filled pauses	0.08	5.03×10^{-7}
mfcc_sma[11]_meanFallingSlope	Mean of the falling slope of the second Mel-frequency cepstral coefficient (MFCC)	0.06	.00069
pcm_fftMag_spectralHarmonicity_sma_risetime	Rise time of the signal for magnitude of psychoacoustic harmonicity	0.05	.0012
mfcc_sma[14]_risetime	Rising time of the second MFCC	0.05	.00096
pcm_fftMag_spectralRollOff90.0_sma_de_minPos	Absolute position of the minimum value of the deltas of magnitude of spectral roll off point 90%	0.05	.00026
mfcc_sma_de[9]_upleveltime25	Percentage of time over 25% of the range of variation of the deltas of the nineth MFCC	0.05	.00064
audSpec_Rfilt_sma[25]_quartile1	First quartile of the RASTA-style filtered auditory spectrum, band 25	0.04	.0023
mfcc_sma[1]_segLenStddev	Standard deviation of the segment lengths of first MFCC	0.04	.0021
audSpec_Rfilt_sma_de[5]_iqr2-3	Interquartile 2-3 of the deltas of the RASTA-style filtered auditory spectrum, band 5	0.04	.00011
pcm_fftMag_fband250-650_sma_de_stddev	Standard deviation of the delta of magnitude of frequency band 250-650 Hz.	0.04	.0024
mfcc_sma_de[2]_lpc1	Linear prediction coefficient one of the deltas	0.04	.0015

	of the second MFCC		
pcm_fftMag_fband250-650_sma_de_rqmean	Root-quadratic mean of the deltas of magnitude of frequency band 250-650 Hz	0.04	.0024
audSpec_Rfilt_sma[7]_upleveltime75	Percentage of time over 75% of the range of variation of the RASTA-style filtered auditory spectrum, band 7	0.03	.0012
mfcc_sma[2]_maxSegLen	Maximum of the segment lengths of the second MFCC	0.03	.0019
audSpec_Rfilt_sma_de[5]_upleveltime75	Percentage of time over 75% of the range of variation of the deltas of RASTA-style filtered auditory spectrum, band 5	0.03	.0023
audSpec_Rfilt_sma_de[5]_upleveltime90	Percentage of time over 90% of the range of variation of the deltas of RASTA-style filtered auditory spectrum, band 5	0.03	.002
audSpec_Rfilt_sma_de[7]_upleveltime75	Percentage of time over 75% of the range of variation of the deltas of RASTA-style filtered auditory spectrum, band 7	0.03	.002
audSpec_Rfilt_sma_de[15]_lpc0	Linear prediction coefficient zero of the delta of RASTA-style filtered auditory spectrum, band 15	0.03	.0016
audSpec_Rfilt_sma_de[15]_lpc1	Linear prediction coefficient one of the deltas of RASTA-style filtered auditory spectrum, band 15	0.03	.00034
audSpec_Rfilt_sma_de[15]_lpc2	Linear prediction coefficient two of the delta of RASTA-style filtered auditory spectrum, band 15	0.03	.00055
audSpec_Rfilt_sma[18]_qregc1	Quadratic regression coefficient 1 of RASTA-style filtered auditory spectrum, band 19	0.03	.00095
audSpec_Rfilt_sma[18]_qregc2	Quadratic regression coefficient 2 of RASTA-style filtered auditory spectrum, band 19	0.03	.00057
audSpec_Rfilt_sma_de[15]_lpc3	Linear prediction coefficient three of the	0.02	.00097

	delta of RASTA-style filtered auditory spectrum, band 15		
audspec_lengthL1norm_sma_peakRangeAbs	Absolute peak range of sum of auditory spectrum	0.02	.0018
pcm_fftMag_spectralRollOff25.0_sma_pctlrange0-1	Outlier robust signal range 'max-min' represented by the range of the 1% and the 99% percentile from the magnitude of spectral roll off point 25%	0.01	2.75×10^{-14}
mfcc_sma_de[4]_peakMeanRel	Relative peak mean of the delta of the fourth MFCC	0.01	.00086
pcm_fftMag_spectralRollOff75.0_sma_quartile1	First quartile of magnitude of spectral roll-off point 75%	0.00	2.17×10^{-18}
pcm_fftMag_spectralRollOff75.0_sma_quartile3	Third quartile of magnitude of spectral roll-off point 75%	0.00	2.00×10^{-18}

[a]Importance was the impurity-based importance score of each acoustic feature that were computed as the mean of accumulation of the impurity decrease within each tree of the random forest.
[b]The *P* value was calculated using a t-test for each acoustic feature. Only the acoustic features with a *P* value less than .0025 were included in the model.

Comparison of performance across different audio recording lengths

In addition to the optimal model based on whole recordings (1+ hour), we further examined the MCI detection performance of various audio recording lengths. As shown in **Figure 3**, in the case of five-minute audio segments, we identified 21 acoustic features that exhibited significant associations with cognitive status (e.g., $P < .0025$). The random forest model constructed using these 21 features achieved an AUC of 0.79 (95% CI: 0.73-0.86). Similarly, for the 10-minute audio segments, we identified 25 significant acoustic features, and achieved an AUC of 0.81 (95% CI: 0.75-0.87). When using 15-minute audio segments, 17 acoustic features were found to be significantly associated with cognitive status, leading to an AUC of 0.80 (95% CI: 0.75-0.86) from the random forest model. Lastly, in the case of 30-minute audio segments, 17 acoustic features were significantly associated with cognitive status, and the random forest model achieved an AUC of 0.82 (95% CI: 0.76-0.89). The accuracy, sensitivity, and specificity of these

models were presented in **Multimedia Appendix 1**. These metrics were computed based on the means and SDs obtained from using 10-fold cross-validation.

Discussion

This study developed a machine learning pipeline to optimize the detection capability of acoustic features for MCI. We identified 29 acoustic features from 200 FHS participants' voice recordings collected at their NP exams, which yielded an accuracy of 87% in classifying those with normal cognition versus MCI. Our findings highlight the significant potential of acoustic-based features of human speech as an easily collectible and accurate data modality for early ADRD detection.

Detecting ADRD early in the disease course and implementing timely interventions to slow its progression continue to be the primary strategies for addressing this condition. The method developed in this study using acoustic features for MCI monitoring aligns well with this goal. Specifically, despite recent FDA approvals for aducanumab and lecanemab as disease-modifying treatments for ADRD, concerns have emerged about the inclusivity of the trial population and the equitable distribution of benefits to all potential beneficiaries[33]. The acoustic feature-based machine learning approach in this study addresses the limited early detection capability of traditional NP tests for asymptomatic individuals, as well as the challenges associated with the cost and time-consuming nature of CSF and blood-based biomarkers[34]. Speech data collection presents a non-invasive and accessible approach for cognitive health monitoring. This motivates potential future applications where passive voice collection tools, like hearing aids, could be employed to gather such data. Utilizing non-semantic, acoustic features of speech offer practical advantages from the perspective of data privacy and security. Unlike linguistic features, which may raise concerns around individual privacy and confidentiality, acoustic features can be derived without the need for direct access

to sensitive personal information. The analysis based on acoustic features reduces privacy concerns and ensures that confidential data remains protected or unidentifiable during the cognitive monitoring process.

Studies examining discourse patterns in participants with ADRD have consistently observed difficulties in word retrieval, less efficient speech, and a notable increase in both the frequency and duration of pauses when their speech is compared to that of healthy adults[35, 36]. Notably, in this study, among the features considered crucial for model performance, those related to filled pauses, such as nrFP and tFP, played a significant role. Filled pauses, such as "um" or "er," are non-lexical vocalizations. In individuals with dementia, pauses in speech are frequently longer and more frequent, which may indicate challenges with semantic and lexical decision-making, cognitive load, and familiarity with topics[36, 37]. This study further highlights that pausing in the speech of individuals with dementia is often considered a dysfluency, serving as a behavioral hallmark that may signify difficulties in social interactions[38]. Our findings are also consistent with previous studies that have examined acoustic-based speech markers in older adults and have found good predictive accuracy in identifying those with MCI as compared to being cognitively normal[39, 40]. Other studies have also found temporal parameters, including prosodic rate and spectrum features, such as Mel-frequency cepstral coefficients (MFCCs), to predict those with MCI or early ADRD[41, 42]. These findings offer a research target for further understanding speech issues and mechanisms related to cognitive health. By integrating acoustic analysis into routine clinical assessments, we can potentially enhance current diagnostic tools. This integration provides clinicians with additional quantitative data to support their diagnostic decisions and monitoring of disease progression. Furthermore, the acoustic features identified in this study hold promise for their potential application in large-scale screening programs aimed at identifying individuals at risk of developing MCI. Such screening tools, leveraging these

features, could offer a cost-effective and scalable approach, enabling broader population reach and early intervention strategies. Thus, these findings not only contribute to our scientific understanding but also have practical implications for improving early detection of cognitive impairment.

A unique contribution of our study that has not been well-examined in previous studies is the impact of the speech recording duration on the model performance. Although the full recording yielded the highest AUC (87%), we did not observe substantial differences in model performance based on varying voice recording lengths (e.g., five minutes, 10 minutes, 15 minutes, and 30 minutes). This finding holds important implications for future studies that involve collecting voice recordings from participants, suggesting that achieving good predictive performance may not require collecting lengthy audio data. It underscores the potential to minimize participant burden and time spent collecting data, while preserving the data's analytical quality. Other strengths of this study include using a community-based sample within a controlled environment for the voice recordings taken during the NP exams. Furthermore, this study employs highly interpretable methods throughout, from feature selection to predictive model construction, achieving good MCI prediction capability. This sets a benchmark for future research attempting more complex analytical approaches. In the future, we can compare complex machine learning methods to fully investigate how to balance the relationship between interpretability and predictive performance.

Important limitations, however, include the inability to account for or investigate the impact of other conditions or risk factors, such as depression[43], that may influence speech patterns within the analysis. Due to the lack of available data on depression at the time of voice recording data collection in FHS, we did not investigate the relationship between depression, cognition, and acoustic features in this study. Future research will be essential to delve into this relationship using more comprehensive cohort datasets. Additionally, our sample consisted

mostly of individuals of being White or of European descent, which could potentially limit the generalizability of our findings to other demographic groups. We also recognize that cognition and MCI are not static entities, and that individuals with MCI can be considered to be cognitively normal at a later point in time[44]. Therefore, it may be possible that some participants were misclassified in terms of their cognitive status in our sample. For example, we acknowledge that the use of NIA-AA criteria[45] offers advantages over the NINCDS-ADRDA and DSM-IV criteria, which were used in this study, to ascertain individuals with MCI since it can provide a more comprehensive and inclusive approach, incorporating multiple pathological features. Additionally, the NIA-AA criteria utilizes objective biomarkers and imaging techniques, enhancing diagnostic accuracy and reproducibility. The voice data used in this study were collected in quiet environments, which to some extent limits the widespread applicability of the study results in different environments, such as in-home settings.

To address these limitations, we plan to expand our research in several ways. First, we aim to include more diverse populations in future studies to assess whether the same acoustic features or different ones yield similar results in distinguishing MCI from normal cognition across various demographic groups. Future research should consider utilizing cohorts with biomarker evidence of neurocognitive disorders for further validation the findings. Additionally, we will explore the inclusion of other medical conditions or factors that may impact model performance, broadening our understanding of how speech patterns can be indicative of cognitive health. Specifically, we recognize that emotions may confound the relationship between speech patterns and cognition. Exploring the detection capability of MCI using voice collected in more real-life environments is another direction for future research. Finally, as we continue to advance in the development of speech-based screening and diagnostic tools, it is crucial to proactively address privacy and data security concerns. While our focus in this paper is primarily on the technical aspects of acoustic feature analysis for

cognitive assessment, we recognize the importance of considering the broader societal implications of deploying such technologies in open source or free market contexts. Safeguards must be implemented to ensure that individuals' privacy rights are respected and that their data is used responsibly and ethically.

Conclusion

This study demonstrated the potential for accurate identification of MCI using non-semantic, acoustic speech features. Our research benefits from a well-defined sample and comprehensive speech data collected during NP exams, which have been rigorously analyzed.

Acknowledgments

We acknowledge FHS participants for their dedication. This study would not be possible without them. We also thank the researchers in FHS for their efforts over the years in the examination of participants.

Data Availability

The derived acoustic features can be requested through a formal research application to the Framingham Heart Study[46].

Funding

This work was supported by the Framingham Heart Study of the National Heart Lung and Blood Institute of the National Institutes of Health and Boston University School of Medicine. Funding for this work was supported in whole or in part with Federal funds from the National Heart Lung and Blood Institute, Department of Health and Human Services, under Contract No. 75N92019D00031 and contracts N01-HC-25195 and HHSN269201500001I, as well as grants from the National Institute on Aging R01-AG016496, R01-AG008122, R01-AG049810, RF1AG054156, R01-AG062109, and U19AG068753. Funding for the analysis of this study came from the National Institute on Aging, grant number R41-AG080977, which supported Brian

Bischoff (main PI), Adrian Lister, Honghuang Lin, Phillip Hwang (subaward PI), Huitong Ding, and Cody Karjadi. Rhoda Au reports conflicts including Signant Health, Biogen, NovoNordisk, and the Davos Alzheimer's Collaborative.

Multimedia Appendix 1

Performance of models for MCI prediction using different audio length segments.

References

1. 2021 Alzheimer's disease facts and figures. Alzheimers Dement. 2021 Mar;17(3):327-406. PMID: 33756057. doi: 10.1002/alz.12328.

2. Gustavsson A, Norton N, Fast T, Frölich L, Georges J, Holzapfel D, et al. Global estimates on the number of persons across the Alzheimer's disease continuum. Alzheimer's & Dementia. 2023;19(2):658-70.

3. Desai AK, Grossberg GT. Diagnosis and treatment of Alzheimer's disease. Neurology. 2005;64(12 suppl 3):S34-S9.

4. Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, et al. Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's & dementia. 2011;7(3):280-92.

5. Tarawneh R, Holtzman DM. The clinical problem of symptomatic Alzheimer disease and mild cognitive impairment. Cold Spring Harbor perspectives in medicine. 2012;2(5).

6. Leifer BP. Early diagnosis of Alzheimer's disease: clinical and economic benefits. Journal of the American Geriatrics Society. 2003;51(5s2):S281-S8.

7. Laske C, Sohrabi HR, Frost SM, López-de-Ipiña K, Garrard P, Buscema M, et al. Innovative diagnostic tools for early detection of Alzheimer's disease. Alzheimer's & Dementia.

2015;11(5):561-78.

8. Robinson P. The cognitive hypothesis, task design, and adult task-based language learning. 2003.

9. Cuetos F, Arango-Lasprilla JC, Uribe C, Valencia C, Lopera F. Linguistic changes in verbal expression: a preclinical marker of Alzheimer's disease. Journal of the International Neuropsychological Society. 2007;13(3):433-9.

10. Deramecourt V, Lebert F, Debachy B, Mackowiak-Cordoliani M, Bombois S, Kerdraon O, et al. Prediction of pathology in primary progressive language and speech disorders. Neurology. 2010;74(1):42-9.

11. Ding H, Mandapati A, Karjadi C, Ang TFA, Lu S, Miao X, et al. Association Between Acoustic Features and Neuropsychological Test Performance in the Framingham Heart Study: Observational Study. Journal of Medical Internet Research. 2022;24(12):e42886.

12. Ding H, Hamel A, Karjadi C, Ang TFA, Lu S, Thomas RJ, et al. Association Between Acoustic Features and Brain Volumes: the Framingham Heart Study. Frontiers in Dementia.2:1214940.

13. Nagumo R, Zhang Y, Ogawa Y, Hosokawa M, Abe K, Ukeda T, et al. Automatic detection of cognitive impairments through acoustic analysis of speech. Current Alzheimer Research. 2020;17(1):60-8.

14. Meghanani A, Anoop C, Ramakrishnan A, editors. An exploration of log-mel spectrogram and MFCC features for Alzheimer’s dementia recognition from spontaneous speech. 2021 IEEE spoken language technology workshop (SLT); 2021: IEEE.

15. Xue C, Karjadi C, Paschalidis IC, Au R, Kolachalama VB. Detection of dementia on voice recordings using deep learning: a Framingham Heart Study. Alzheimer's research & therapy. 2021;13:1-15.

16. Vincze V, Szatlóczki G, Tóth L, Gosztolya G, Pákáski M, Hoffmann I, et al. Telltale silence:

temporal speech parameters discriminate between prodromal dementia and mild Alzheimer's disease. Clinical Linguistics & Phonetics. 2021;35(8):727-42.

17. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology. 1984;34(7):939-.

18. Satizabal CL, Beiser AS, Chouraki V, Chêne G, Dufouil C, Seshadri S. Incidence of dementia over three decades in the Framingham Heart Study. New England Journal of Medicine. 2016;374(6):523-32.

19. Anguera X, Bozonnet S, Evans N, Fredouille C, Friedland G, Vinyals O. Speaker diarization: A review of recent research. IEEE Transactions on audio, speech, and language processing. 2012;20(2):356-70.

20. Bredin H, Laurent A. End-to-end speaker segmentation for overlap-aware resegmentation. arXiv preprint arXiv:210404045. 2021.

21. Bredin H, Yin R, Coria JM, Gelly G, Korshunov P, Lavechin M, et al., editors. Pyannote. audio: neural building blocks for speaker diarization. ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2020: IEEE.

22. Eyben F, Wöllmer M, Schuller B, editors. Opensmile: the munich versatile and fast open-source audio feature extractor. Proceedings of the 18th ACM international conference on Multimedia; 2010.

23. Boersma P. Praat: doing phonetics by computer [Computer program]. http://www praat org/. 2011.

24. Schuller B, Steidl S, Batliner A, Vinciarelli A, Scherer K, Ringeval F, et al., editors. The INTERSPEECH 2013 computational paralinguistics challenge: Social signals, conflict, emotion, autism. Proceedings INTERSPEECH 2013, 14th Annual Conference of the International Speech

Communication Association, Lyon, France; 2013.

25. Dumpala SH, Rodriguez S, Rempel S, Sajjadian M, Uher R, Oore S. Detecting depression with a temporal context of speaker embeddings. Proc AAAI SAS. 2022.

26. Luz S, Haider F, de la Fuente Garcia S, Fromm D, MacWhinney B. Alzheimer's dementia recognition through spontaneous speech. Frontiers in computer science. 2021;3:780169.

27. Beccaria F, Gagliardi G, Kokkinakis D, editors. Extraction and Classification of Acoustic Features from Italian Speaking Children with Autism Spectrum Disorders. Proceedings of the RaPID Workshop-Resources and ProcessIng of linguistic, para-linguistic and extra-linguistic Data from people with various forms of cognitive/psychiatric/developmental impairments-within the 13th Language Resources and Evaluation Conference; 2022.

28. Sümer Ö, Beyan C, Ruth F, Kramer O, Trautwein U, Kasneci E. Estimating Presentation Competence using Multimodal Nonverbal Behavioral Cues. arXiv preprint arXiv:210502636. 2021.

29. Sidorov M, Ultes S, Schmitt A, editors. Automatic recognition of personality traits: A multimodal approach. Proceedings of the 2014 Workshop on Mapping Personality Traits Challenge and Workshop; 2014.

30. Weninger F, Eyben F, Schuller BW, Mortillaro M, Scherer KR. On the acoustics of emotion in audio: what speech, music, and sound have in common. Frontiers in psychology. 2013;4:292.

31. de Jong NH, Pacilly J, Heeren W. PRAAT scripts to measure speed fluency and breakdown fluency in speech automatically. Assessment in Education: Principles, Policy & Practice. 2021;28(4):456-76.

32. Breiman L. Random forests. Machine learning. 2001;45:5-32.

33. Manly JJ, Glymour MM. What the aducanumab approval reveals about Alzheimer disease research. JAMA neurology. 2021;78(11):1305-6.

34. Dokholyan NV, Mohs RC, Bateman RJ. Challenges and progress in research, diagnostics,

and therapeutics in Alzheimer's disease and related dementias. Alzheimer's & Dementia: Translational Research & Clinical Interventions. 2022;8(1):e12330.

35. König A, Satt A, Sorin A, Hoory R, Toledo-Ronen O, Derreumaux A, et al. Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring. 2015;1(1):112-24.

36. Pistono A, Pariente J, Bézy C, Lemesle B, Le Men J, Jucla M. What happens when nothing happens? An investigation of pauses as a compensatory mechanism in early Alzheimer's disease. Neuropsychologia. 2019;124:133-43.

37. Merlo S, Mansur LL. Descriptive discourse: topic familiarity and disfluencies. Journal of Communication Disorders. 2004;37(6):489-503.

38. Davis BH, Maclagan M. Examining pauses in Alzheimer's discourse. American Journal of Alzheimer's Disease & Other Dementias®. 2009;24(2):141-54.

39. Kato S, Homma A, Sakuma T. Easy screening for mild Alzheimer's disease and mild cognitive impairment from elderly speech. Current Alzheimer Research. 2018;15(2):104-10.

40. Themistocleous C, Eckerström M, Kokkinakis D. Identification of mild cognitive impairment from speech in Swedish using deep sequential neural networks. Frontiers in neurology. 2018;9:975.

41. Tóth L, Hoffmann I, Gosztolya G, Vincze V, Szatlóczki G, Bánréti Z, et al. A speech recognition-based solution for the automatic detection of mild cognitive impairment from spontaneous speech. Current Alzheimer Research. 2018;15(2):130-8.

42. López-de-Ipiña K, Alonso J-B, Travieso CM, Solé-Casals J, Egiraun H, Faundez-Zanuy M, et al. On the selection of non-invasive methods based on speech analysis oriented to automatic Alzheimer disease diagnosis. Sensors. 2013;13(5):6730-45.

43. Cannizzaro M, Harel B, Reilly N, Chappell P, Snyder PJ. Voice acoustical measurement of the severity of major depression. Brain and cognition. 2004;56(1):30-5.

44. Koepsell TD, Monsell SE. Reversion from mild cognitive impairment to normal or near-normal cognition: risk factors and prognosis. Neurology. 2012;79(15):1591-8.

45. Jack Jr CR, Albert M, Knopman DS, McKhann GM, Sperling RA, Carillo M, et al. Introduction to revised criteria for the diagnosis of Alzheimer's disease: National Institute on Aging and the Alzheimer Association Workgroups. Alzheimer's & dementia: the journal of the Alzheimer's Association. 2011;7(3):257.

46. Framingham Heart Study For Researchers. 2024 [cited 2024 04/22/2024]; Available from: https://www.framinghamheartstudy.org/fhs-for-researchers/.

Figure 1. The machine learning pipeline for MCI detection from voice recordings.

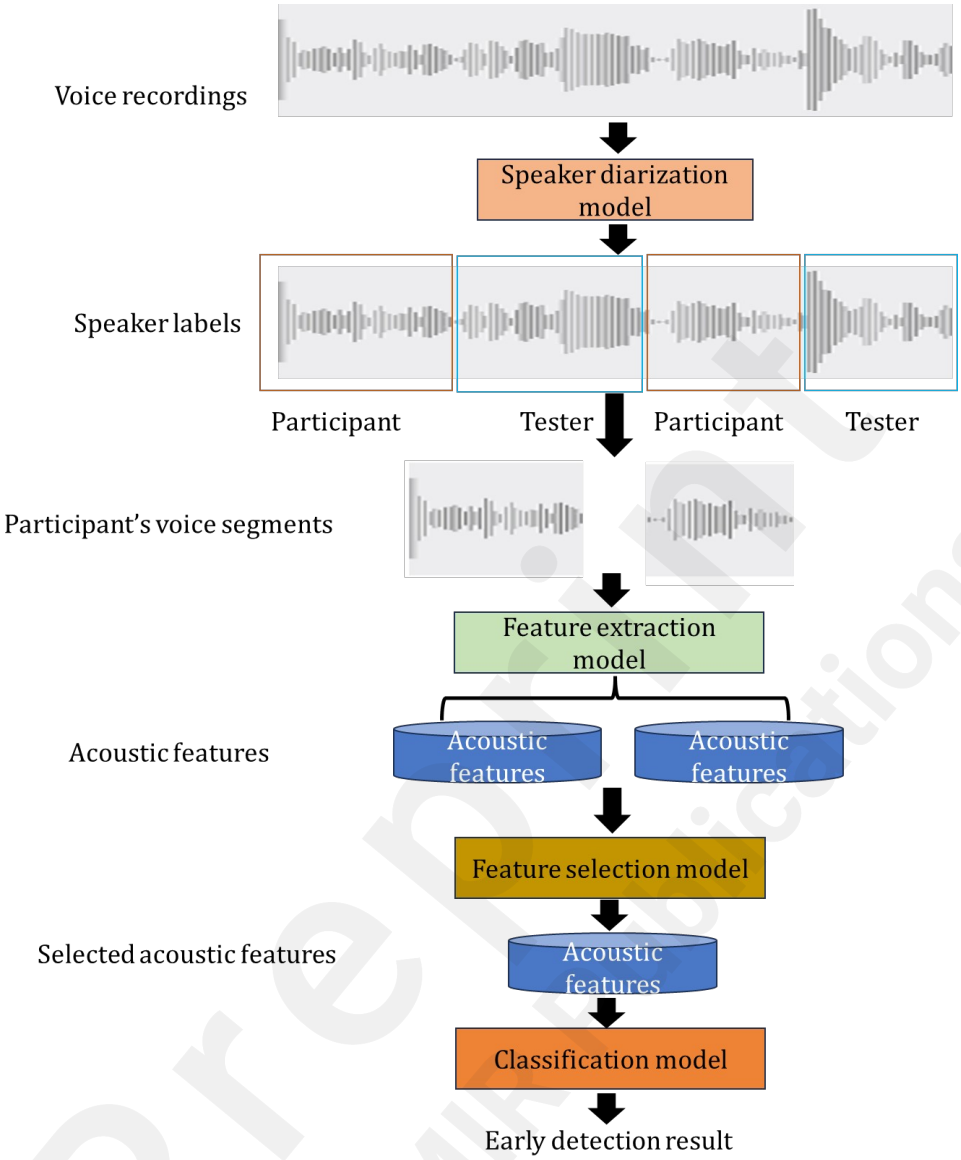


Figure 2. Receiver operating characteristic (ROC) curve of the random forest model for MCI classification. The mean ROC is depicted by the blue line, while the shaded grey area surrounding the curve represents confidence intervals, offering insights into the associated uncertainty of the curve.

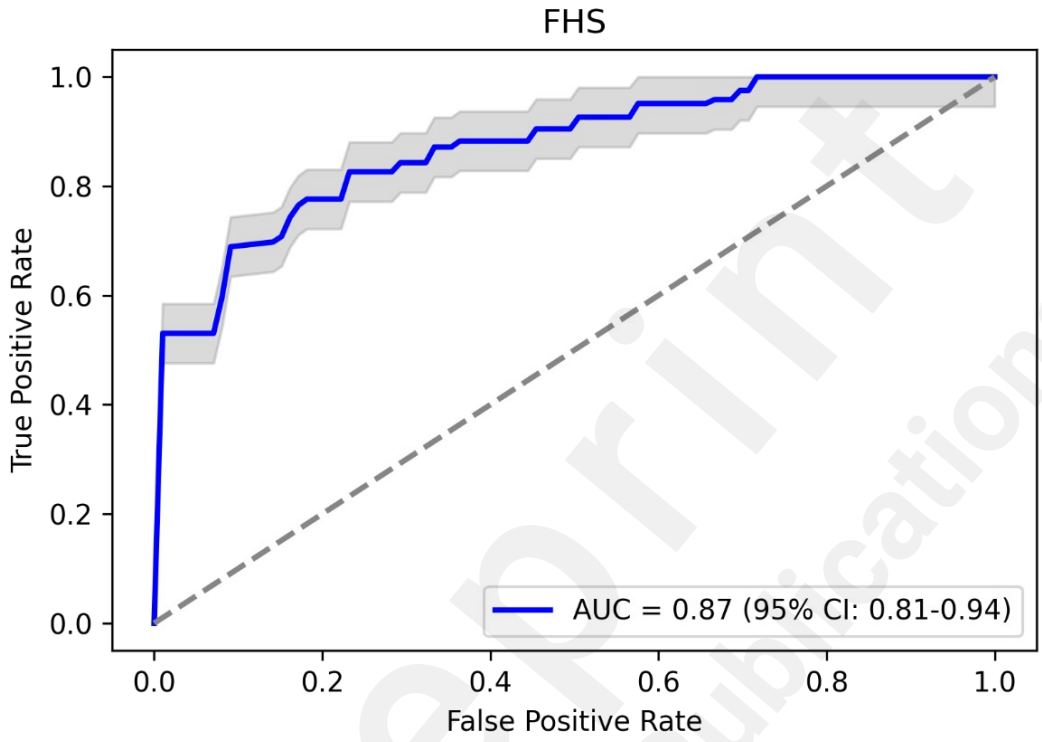
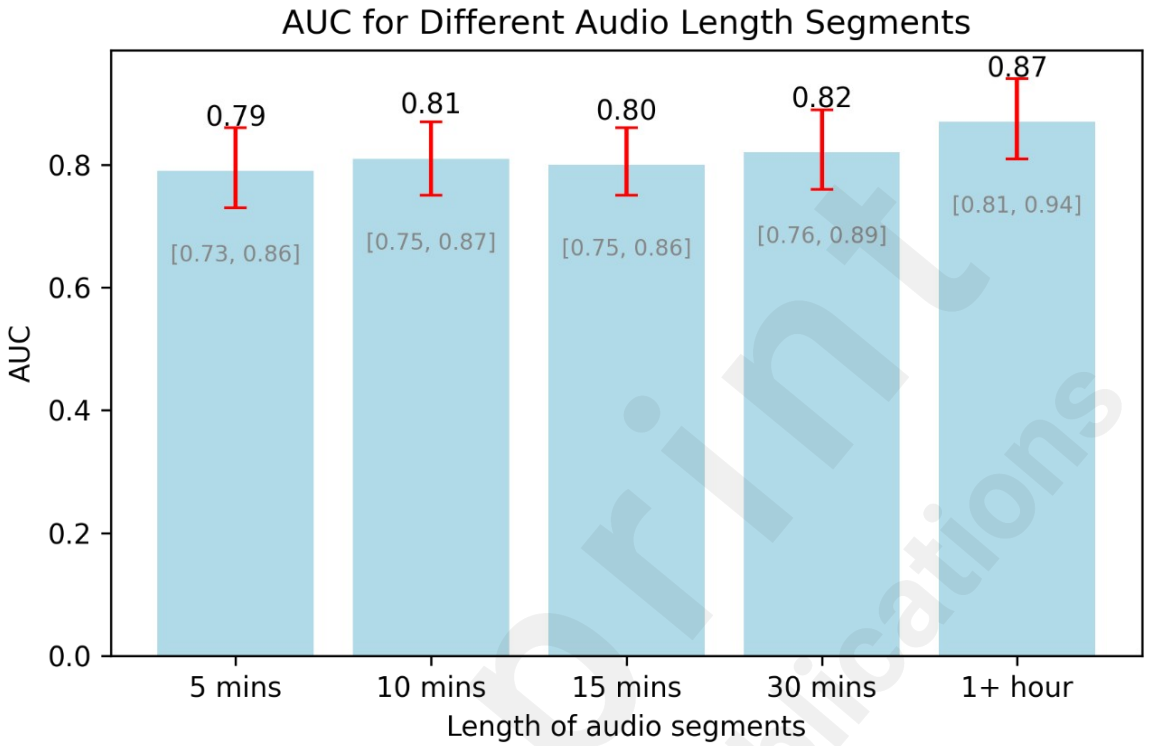


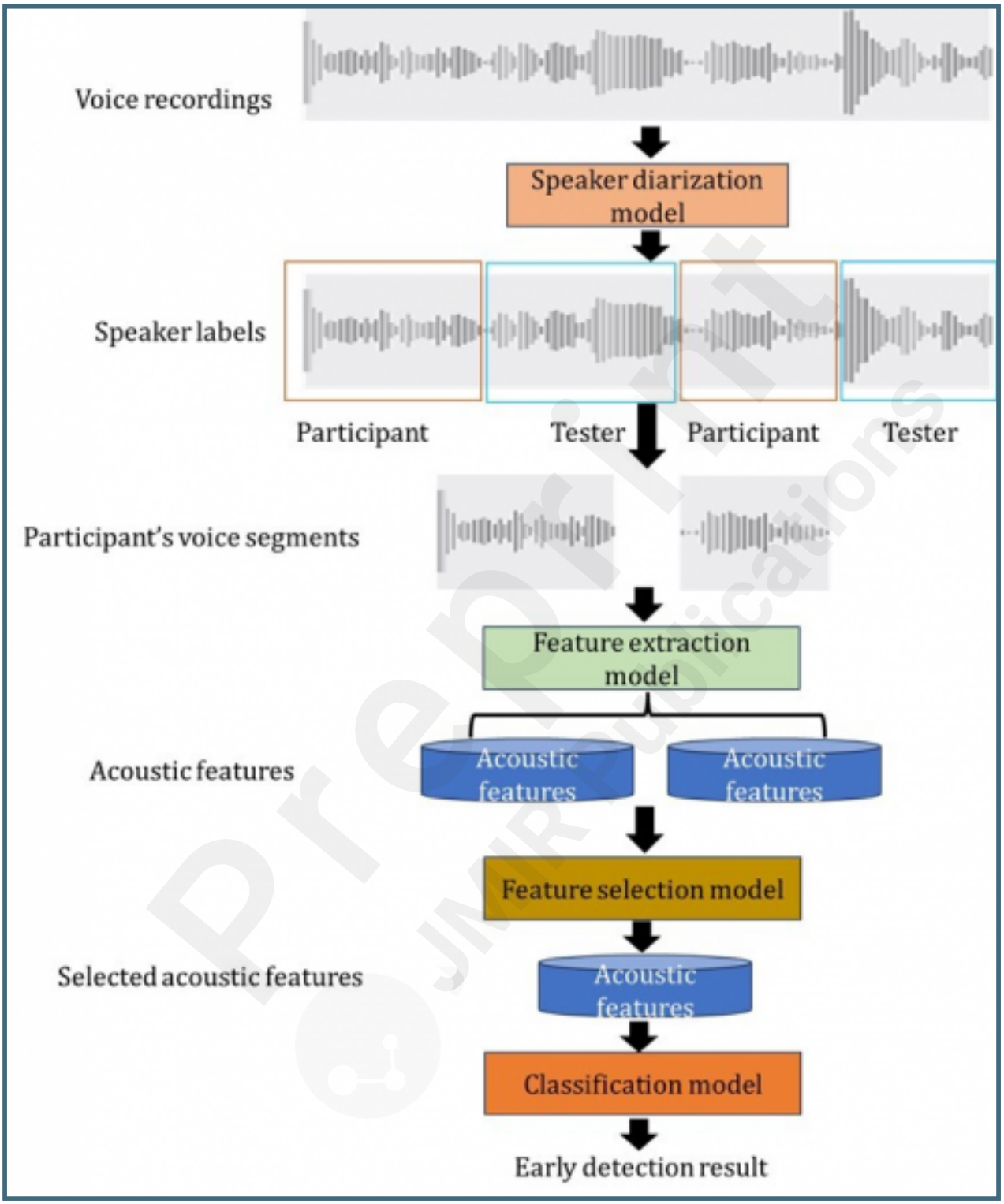
Figure 3. AUCs for MCI detection using different audio length segments. The bar chart depicts the mean AUC of the MCI detection model using various audio length segments, with confidence intervals indicated by the red line.



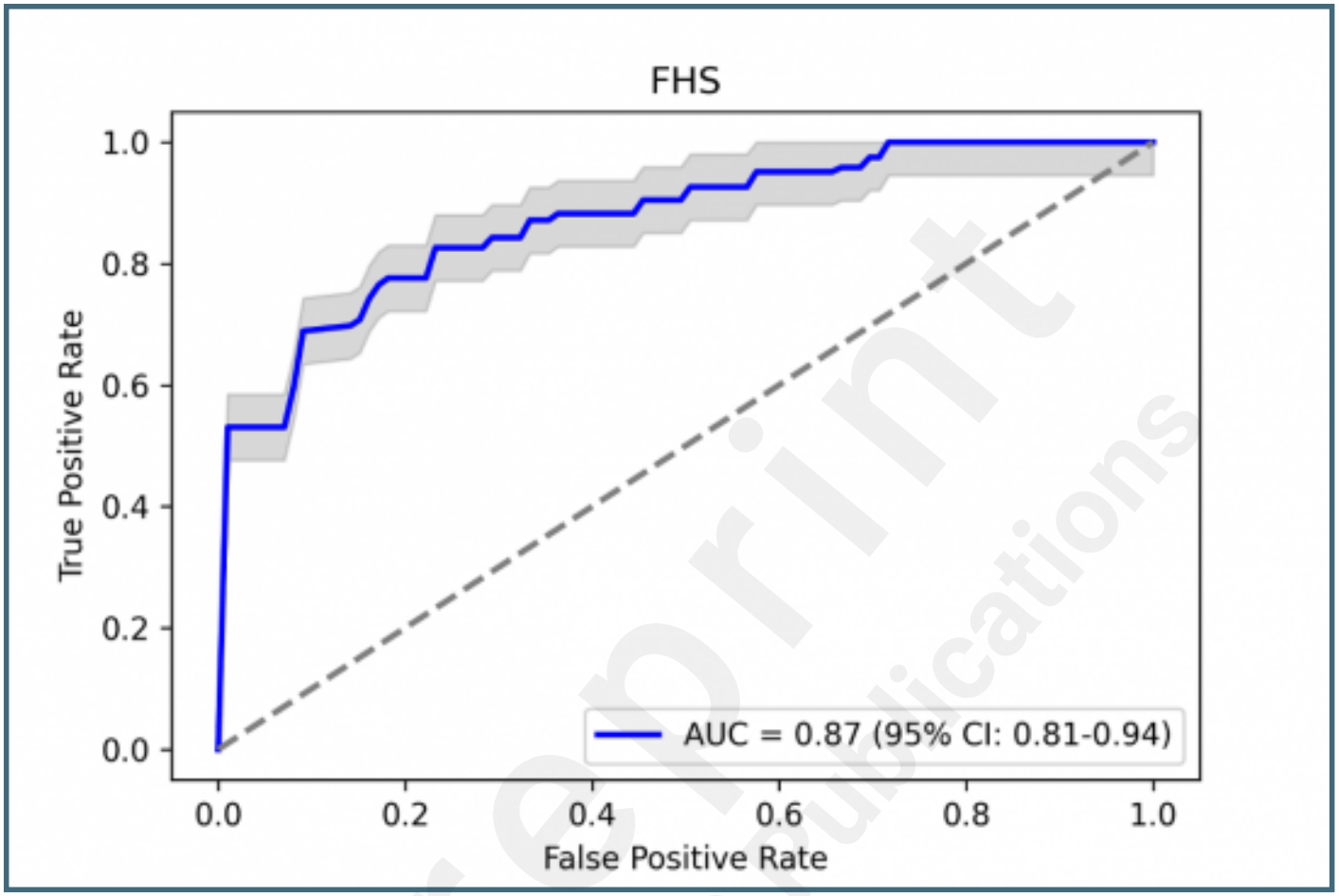
Supplementary Files

Figures

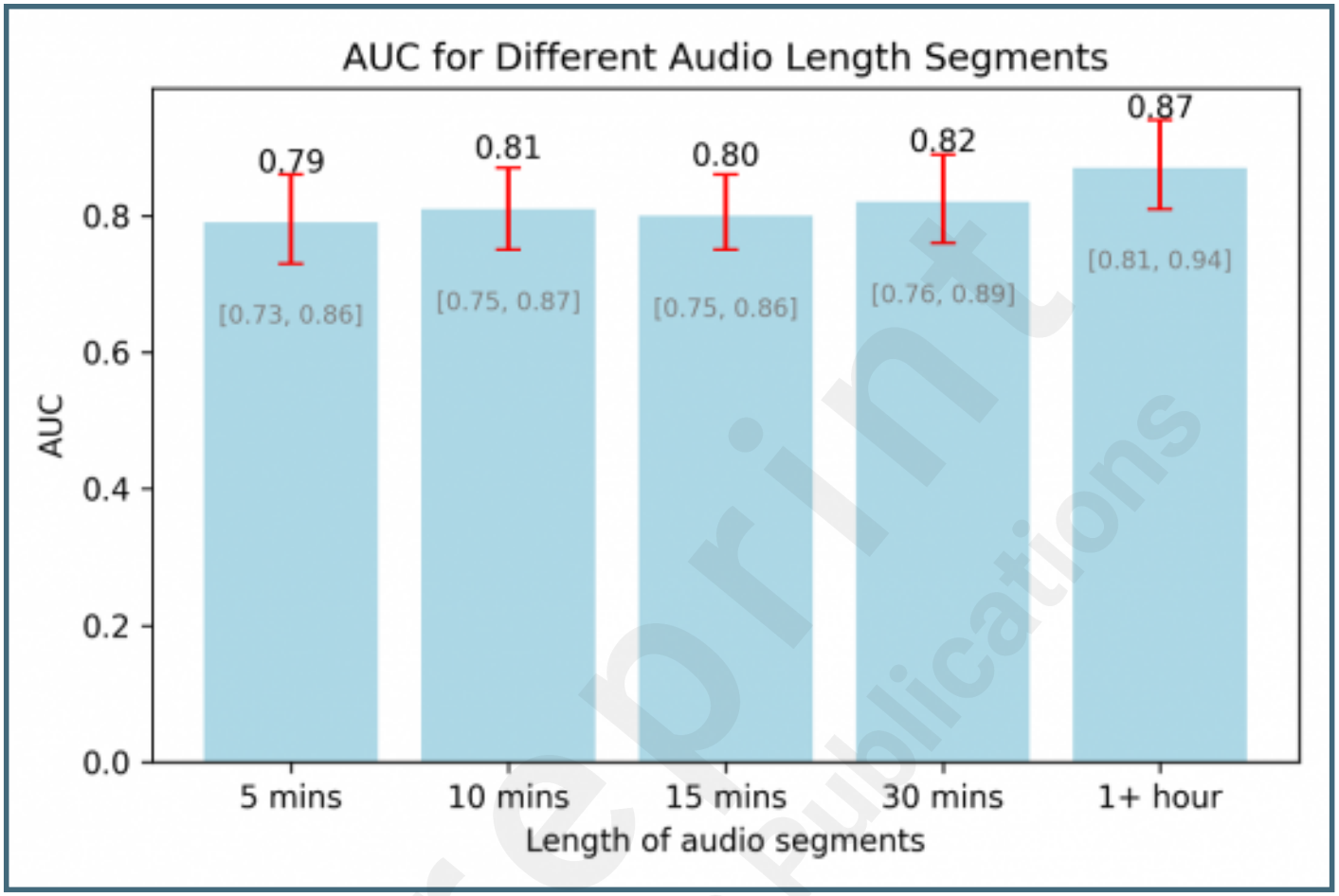
The machine learning pipeline for MCI detection from voice recordings.



Receiver operating characteristic (ROC) curve of the random forest model for MCI classification. The mean ROC is depicted by the blue line, while the shaded grey area surrounding the curve represents confidence intervals, offering insights into the associated uncertainty of the curve.



AUCs for MCI detection using different audio length segments. The bar chart depicts the mean AUC of the MCI detection model using various audio length segments, with confidence intervals indicated by the red line.



Multimedia Appendixes

Table S1.
URL: http://asset.jmir.pub/assets/c396db0bcdbcf7624a668479c09e7ca7.docx