Developing a Multimodal Screening Algorithm for Mild Cognitive Impairment and Early Dementia in Home Healthcare: Protocol for a Cross-Sectional Case–Control Study Using Speech Analysis, Large Language Models, and Electronic Health Records

Maryam Zolnoori

Submitted to: JMIR Research Protocols
on: August 20, 2025

Table of Contents

Developing a Multimodal Screening Algorithm for Mild Cognitive Impairment and Early Dementia in Home Healthcare: Protocol for a Cross-Sectional Case–Control Study Using Speech Analysis, Large Language Models, and Electronic Health Records

Maryam Zolnoori[1] PhD

[1] Columbia University Irving Medical Center New York US

Corresponding Author:
Maryam Zolnoori PhD

Columbia University Irving Medical Center
560 W 168th St
New York
US

Abstract

Background: Mild cognitive impairment and early dementia (MCI-ED) are a growing public health challenge, yet more than half of affected patients remain undiagnosed, particularly in home healthcare (HHC). Traditional biomarkers are costly or invasive, and existing federally mandated assessments such as the Outcome and Assessment Information Set (OASIS) lack sensitivity to subtle cognitive decline. Spoken language and interactional cues are among the earliest indicators of cognitive impairment and, when integrated with routinely collected electronic health record (EHR) data, may enable earlier and more equitable detection.

Objective: This protocol describes the development and evaluation of a multimodal screening algorithm for timely identification of MCI-ED in HHC. The approach leverages three complementary aims: (1) modeling speech and interactional dynamics from patient–nurse encounters, (2) adapting large language models (LLMs) to extract MCI-ED–related information from HHC notes and transcripts, and (3) developing a multimodal screening algorithm that integrates speech, LLM-extracted constructs, and OASIS data.

Methods: We conduct a cross-sectional case–control study in collaboration with VNS Health among patients aged ?65 years. Each participant contributes three to four audio-recorded encounters with nurses, yielding approximately 120 minutes of spontaneous speech, which are linked to OASIS assessments and EHR notes.
• Aim 1: Speech and interaction modeling. Acoustic, linguistic, emotion, and interactional features are extracted using automated pipelines. Measures include phonetic motor planning, lexical richness, semantic fluency, syntactic organization, emotional expression, and patient–nurse interactional dynamics (e.g., turns, timing, interactivity).
• Aim 2: LLM-based extraction. We define an ontology for clinical symptoms, lifestyle risk factors, and communication deficits, and adapt Llama-3.1 models in a secure environment. Hybrid strategies (prompted extraction and parameter-efficient fine-tuning) extract and normalize concepts from HHC notes and transcripts. Outputs are aggregated at the patient level.
• Aim 3: Multimodal algorithm. Features from Aims 1–2 are fused with OASIS data. Multiple classifiers, including support vector machines, logistic regression, ensemble trees, and deep neural models, are evaluated using stratified nested cross-validation. Model performance is assessed using AUROC, AUPRC, calibration, and fairness metrics (equality of opportunity/odds) across sex, race, and age groups. The target AUROC is >0.85 across subgroups.

Results: We enroll 47 HHC patients (51.1% female; 55.3% Black, 29.8% White, 8.5% Asian, 6.4% Hispanic). Encounters average 19 minutes, with nurses contributing more words than patients (median 842 vs. 589). Preliminary analyses show that integrating speech with EHR/OASIS outperforms single-source models. Recruitment and validation are ongoing.

Conclusions: This protocol demonstrates the feasibility of embedding multimodal MCI-ED screening in HHC. Early results highlight the potential of integrating spontaneous speech, interactional dynamics, and routinely collected clinical data to improve timely identification of cognitive decline. Definitive evaluation of algorithm performance and fairness follows full accrual.

(JMIR Preprints 20/08/2025:82731)
DOI: https://doi.org/10.2196/preprints.82731

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
- No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in <a href="https:/

Original Manuscript

Developing a Multimodal Screening Algorithm for Mild Cognitive Impairment and Early Dementia in Home Healthcare: Protocol for a Cross-Sectional Case–Control Study Using Speech Analysis, Large Language Models, and Electronic Health Records

Maryam Zolnoori, PhD[1,2]
[1]Columbia University Irving Medical Center, New York, NY
[2]School of Nursing, Columbia University, New York, NY 10032, United States

**Authors contributed equally*

Corresponding Author:
Maryam Zolnoori, PhD
Columbia University Medical Center,
School of Nursing Columbia University
560 W 168th St, New York, NY 10032, United States
USA Phone: 317-515-1950
E-mail: mz2825@cumc.columbia.edu; m.zolnoori@gmail.com

Conflict of Interest: The authors declare no conflicts of interest.

Funding Acknowledgment: - K99/R00AG076808 "Development of a Screening Algorithm for Timely Identification of Patients with Mild Cognitive Impairment and Early Dementia in Home Healthcare" from National Institute on Aging;

Pease Note that the reserch methodoly provided in this manuscript is from K99/R00AG076808 that was previously reviewed by NIA and received score 18 and funded.

ABSTRACT

Background:
Mild cognitive impairment and early dementia (MCI-ED) are a growing public health challenge, yet more than half of affected patients remain undiagnosed, particularly in home healthcare (HHC). Traditional biomarkers are costly or invasive, and existing federally mandated assessments such as the Outcome and Assessment Information Set (OASIS) lack sensitivity to subtle cognitive decline. Spoken language and interactional cues are among the earliest indicators of cognitive impairment and, when integrated with routinely collected electronic health record (EHR) data, may enable earlier and more equitable detection.

Objective:
This protocol describes the development and evaluation of a multimodal screening algorithm for timely identification of MCI-ED in HHC. The approach leverages three complementary aims: (1) modeling speech and interactional dynamics from patient–nurse encounters, (2) adapting large language models (LLMs) to extract MCI-ED–related information from HHC notes and transcripts, and (3) developing a multimodal screening algorithm that integrates speech, LLM-extracted constructs, and OASIS data.

Methods:
We conduct a cross-sectional case–control study in collaboration with VNS Health among patients aged ≥65 years. Each participant contributes three to four audio-recorded encounters with nurses, yielding approximately 120 minutes of spontaneous speech, which are linked to OASIS assessments and EHR notes.

- **Aim 1: Speech and interaction modeling.** Acoustic, linguistic, emotion, and interactional features are extracted using automated pipelines. Measures include phonetic motor planning, lexical richness, semantic fluency, syntactic organization, emotional expression, and patient–nurse interactional dynamics (e.g., turns, timing, interactivity).
- **Aim 2: LLM-based extraction.** We define an ontology for clinical symptoms, lifestyle risk factors, and communication deficits, and adapt Llama-3.1 models in a secure environment. Hybrid strategies (prompted extraction and parameter-efficient fine-tuning) extract and normalize concepts from HHC notes and transcripts. Outputs are aggregated at the patient level.
- **Aim 3: Multimodal algorithm.** Features from Aims 1–2 are fused with OASIS data. Multiple classifiers, including support vector machines, logistic regression, ensemble trees, and deep neural models, are evaluated using stratified nested cross-validation. Model performance is assessed using AUROC, AUPRC, calibration, and fairness metrics (equality of opportunity/odds) across sex, race, and age groups. The target AUROC is >0.85 across subgroups.

Results (initial implementation):
We enroll 47 HHC patients (51.1% female; 55.3% Black, 29.8% White, 8.5% Asian, 6.4% Hispanic). Encounters average 19 minutes, with nurses contributing more words than patients (median 842 vs. 589). Preliminary analyses show that integrating speech with EHR/OASIS outperforms single-source models. Recruitment and validation are ongoing.

Conclusions:

This protocol demonstrates the feasibility of embedding multimodal MCI-ED screening in HHC. Early results highlight the potential of integrating spontaneous speech, interactional dynamics, and routinely collected clinical data to improve timely identification of cognitive decline. Definitive evaluation of algorithm performance and fairness follows full accrual.

Keywords: mild cognitive impairment; dementia; speech analysis; natural language processing; large language models; home healthcare; screening; multimodal fusion; fairness

INTRODUCTION

Mild cognitive impairment (MCI) and early-stage dementia (ED) represent a looming public health crisis, affecting one-in-five older adults over age 60.[1–3] MCI-ED patients are frequent utilizers of healthcare services in general[4,5,6] and emergency department services[7,8] in particular, and they incur higher costs of care compared with non-MCI patients.[4,9] Despite nationwide efforts for timely diagnosis of MCI, more than 50% of these patients remain underdiagnosed and undertreated.[10–12] This is mostly due to patients' inability to recognize early symptoms,[13] limited availability of biomarkers (e.g., cerebrospinal fluid),[14] and clinicians' insufficient time to assess patients for MCI-ED.[15] Given the projection of 11 to 16 million MCI-ED patients by 2050,[16–19] development of a robust screening tool for early identification of elderly patients with MCI-ED has been recognized as a research priority by the National Institute on Aging (NIA).[20,12] We propose to address barriers to early identification of MCI-ED via the development of an innovative algorithm built on a combination of multiple data streams, including data extracted from electronic health records (EHR) and audio-recorded patient-clinician verbal communication during routine encounters.

Emerging studies have shown that language impairment is one of the earliest signs of cognitive decline, enabling spoken language to act as a biomarker for early diagnosing MCI-ED.[21,22] Features of spoken language can characterize multiple aspects of cognitive dysfunctions including memory impairment, phonetic motor disorder[1,23], language impairment,[21] and disturbance in executive functioning,[24] resulting in a patient's deficit in communication and daily functioning.[1] Given the centrality of language to cognitive functioning, we hypothesize that quantifying properties of patients' verbal communications with clinicians and integrating the properties with patients' clinical histories and lifestyle risk factors extracted from the EHR system can improve the performance of the screening algorithm for timely identification of elderly patients with MCI-ED.

Our preliminary study[2,25] on utilizing natural language processing (NLP) to identify risk factors associated with severe cognitive impairment showed that 63% of home healthcare (HHC) patients with cognitive impairment remained underdiagnosed due to inadequate transfer of information on the HHC patient's cognitive status to HHC clinicians and lack of adequate time and training of the clinicians for timely diagnosis of cognitive impairment. Our primary goal is to utilize the routinely generated data in the HHC setting, including OASIS (Outcome and Assessment Information Set - a federally required assessment of patients admitted to HHC dataset), HHC nurses' notes, and patient-nurse verbal communication to develop a screening algorithm for timely identification of MCI-ED patients. This proposed study was built on our current pilot study that identified a practical approach for audio-recording HHC patient-nurse verbal communication and speech analysis approaches for quantifying properties of patient's language and modeling their interactions with HHC nurses. Specifically, we aim to:

Aim 1: Model MCI-ED patients' verbal communications with clinicians using an automated speech analysis system. Specifically, this system would be able to model MCI-ED patients' abilities in 1) phonetic motor planning, 2) semantic and pragmatic levels of language organization, 3) vocal and semantic expression of emotion, and 4) interaction with clinicians. Achieving this aim will generate a list of speech and communication features that are associated with the patient's cognitive dysfunction and communication deficits and will be used for identifying MCI-ED patients.

Aim 2: Utilize large language model (LLM) to automatically identify MCI-ED related information, including i) clinical symptoms, ii) lifestyle risk factors, and iii) communication deficits from both HHC clinical notes and patient-clinician verbal communication. Achieving this aim will extend the adaptation strategies for LLMs to automatically identify MCI-ED related information from free-text datasets.

Aim 3: Develop a screening algorithm to identify HHC patients with MCI-ED. This screening algorithm will be developed and tested iteratively using: 1) OASIS dataset, 2) HHC clinical notes, and 3) verbal communication features. We hypothesize that adding speech features to the clinical notes and the OASIS dataset will improve the performance of the screening algorithm.

METHODS

Study setting and study population: This study is conducted in collaboration with the VNS Health, one of the largest home healthcare system in the U.S. The population of this study are patients 65 years old and older who receive HHC services from the VNS Health.

Eligibility criteria for the case and control groups: Developing this screening algorithm requires a comparison of the speech patterns and characteristics of MCI-ED patients and non-cognitively impaired patients. For each patient, we audio-record three to four HHC patient-nurse encounters to quantify the temporal parameters of speech components. The study design is a cross-sectional case-control study. The case group includes MCI-ED patients, and the control group includes non-cognitively impaired patients. 1) Inclusion criteria for the case group: (a) we use EHR to identify patients with a clinical diagnosis of MCI-ED (International Classification of Diseases 10 [ICD-10] code: G31.84); (b) since more than 50% of patients with MCI-ED remain underdiagnosed, we use screening tool, the Montreal Cognitive Assessment (MoCA)[26,27] for MCI-ED detection. Patients with MoCA score of 18-25 are highly suspected for presence of MCI-ED. We will notify the patient's primary care physician for referral to a neurologist for further evaluation. 2) Inclusion criteria for the control group: patients with no clinical diagnosis for cognitive impairment will be included in the control group. 2) Exclusion criteria: (a) the patient has limited ability to communicate independently with the HHC nurse (e.g., difficulty with English language) or (b) the patient has speech and language disorders due to neurological conditions (e.g., seizure disorder, Parkinson's disease) other than MCI-ED.

Study design and sample size: (i) Sample size for the development dataset: The development dataset is used for developing and fine-tuning ML models (Aim 3) to build the screening algorithm. We used the sample size formula specified by Charan et al.[28] To detect a medium effect size (Cohen's d=0.50) using two tailed alpha level of 0.05 with 80% power, we would need a sample size of ≈ 100 HHC patients for case and control groups (N=200). To ensure validity and reliability of the algorithm in identification of MCI-ED patients for a racially diverse population, we recruit an equal sample of patients from each racial group. At this stage, including other races/ethnic groups was not feasible due to the required large sample size and recruitment considerations. (ii) Sample size for validation (test) dataset: We create a validation dataset to generate an unbiased estimate of the performance of ML models when comparing or selecting between final models. We use a sample size of 40 HHC patients with MCI-ED and 40 non-cognitively impaired HHC patients that include equal numbers of each race (non-Hispanic White and non-Hispanic Black groups).

Audio-recording patient-nurse verbal communication: If the nurse consents to participate, the research assistant attends the patient-nurse encounter and operate the Saramonic Blink device to audio-record the encounter.

EHRs: We link the audio-recorded patient-nurse verbal communication for the sample of this study to the OASIS dataset, supplemental EHR items (e.g., medications), and free-text clinical notes extracted from VNSNY's EHR system. OASIS is the most comprehensive data on the HHC assessment and outcomes, including information about patients' health, functional status, and living arrangements.[29,30] Clinical free text notes include two types of entries. *Visit notes* contain the nurse's narrative at each HHC visit, structured as one record per visit. *Coordination of care* notes document communications the nurse may have at any time with other HHC clinicians, the patient's physician, and family.

Analytic methods

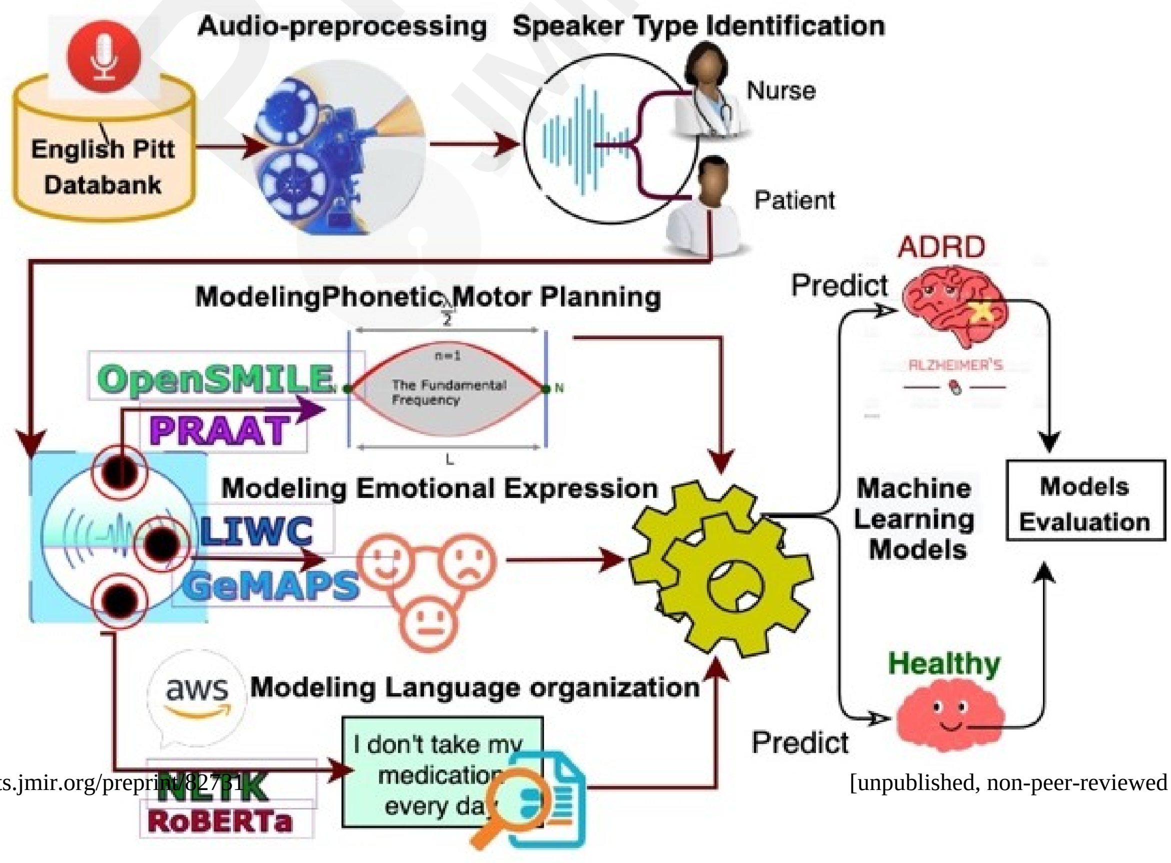
(1) Feasibility study to identify best practical approach for audio-recording patient-nurse verbal interaction in the HHC setting.[31] The study that was conducted at VNS Health, first, we evaluated the functionality and usability of seven audio-recording devices in a laboratory setting. Three devices, Saramonic Blink65[32], Sony ICD-TX6, and Black Vox 365 were selected for further evaluation in the HHC setting by participation of two nurses who audio-recorded patient-nurse verbal communication. Nurses completed the System Usability Scale (SUS) questionnaire to report their feedback about each device. We also evaluated the accuracy of automatic transcription of recorded communications for the devices using the Amazon Web Service (AWS) Transcribe that was measured by Word Error Rate (WER) metric. Saramonic Blink received the best overall evaluation of SUS= 65% and WER = 26%. Second, we conducted semi-structured interviews with three HHC nurses and ten HHC patients to understand facilitators and barriers of audio-recording HHC encounters. Overall, patients found the process of audio-recording satisfactory and convenient with minimal perceived impact on their communications with nurses. Nurses also found the audio-recording process easy to learn and satisfactory. Implications for current study: This pilot study established the feasibility and acceptability of audio-recording HHC patient-nurse encounters. To minimize the risk of increasing the nurses' workload, our research assistant will attend the encounters to operate Saramonic Blink and store the recorded communications in a secure VNSNY server.

(2) Using ML to distinguish between speaker types in recorded HHC nurse-patient verbal communication.[33] Based on the results of the pilot feasibility study, patient-nurse verbal communications were recorded using Saramonic Blink for 10 HHC patient-nurse encounters. In phase (i) of this study, we pre-processed audio-recoded communications using the iZOTOPE-RX 8[34] toolkit to remove background and stationary noises. In phase (ii), all recorded communications were transcribed using AWS Transcribe that assigns speaker_0 or speaker_1 to each participant's utterance in the communication. An utterance is a continuous block of speech (patient or nurse) without interruption. In total, we had 618 patient utterances and 623 nurse utterances. In phase (iii), we trained different ML algorithms on features generated from the patients' and nurses' utterances using classic word-embedding methods (TF-IDF, Word2vec, Glove) and measured their performance using the K-fold cross validation method. XG-boost had a relatively high performance with Area of Under Curve-Receiver-Operating Characteristic (AUC-ROC = 0.80).

By conducting this study,_we established the feasibility of developing ML models for automated speaker type identification, with relatively high classification performance. The proposed study will further expand the size of the training data which would likely result in improved classification performance.

(3) Development of an analytic pipeline for modeling the spoken language in patients with cognitive impairment.[35] We tested performance of this pipeline on the English Pitt Databank that includes audio-recorded description of 174 participants (87 with ADRD [case group], 87 non-cognitively impaired [control group]) of the “Cookie-Theft” test. This test is a drawing of two children stealing cookies behind their mother's back. Patients’ verbal description of this test was proven effective for assessing cognitive function. This pipeline consists of the following six components (**Figure 1**): (i) Data preprocessing; (ii) Speaker type identification; (iii) Phonetic motor planning of the participant was modeled using acoustic parameters in four domains: speech fluency, frequency and spectral domain, voice instability, and voice intensity (see Table 1, component 1). OpenSMILE[36] and PRAAT Vocal Toolkits[37] were used for computing the parameters; (iv) Emotional expression of the participant was modeled using the Geneva Minimalistic acoustic parameter set (GeMAPS),[38] which was developed to index affective physiological changes in non-verbal vocalization of voice production (Table 1, component 2). The GeMAPS capability for modeling non-verbal vocalization of emotion has been shown in real-world test scenarios.[38,39]Additionally, we modeled semantic expression of the emotion using Linguistic Inquiry and Word Count (LIWC) version 2015.[40] LIWC is a text analytic program built on a dictionary that groups related words (e.g., sadness, depressed) into different categories of emotion (e.g., depression); (v) Syntactic and semantic level of language organization was modeled using language parameters in two domains, lexical richness and the patient’s recall ability (Table 1, component 3) using AWS Transcribe (for automatic transcription of participants’ descriptions, preliminary study D2.1) and the NLTK toolkit.[41] We also used distilled RoBERTa,[42] a robustly optimized version of the BERT language model to represent the contextual relations among words or sub-words in the content of the patient’s description; and (vi) Development and evaluation of ML models. We split the dataset into a development (80%) and validation (20%) dataset. We trained different ML algorithms (e.g., XG-boost, ExtraTrees, SVM) on the development dataset. Stratified 5-fold nested cross validation was used for hyperparameter tuning and mitigating the risk of models over-fitting. We measured performance of the ML algorithms using the AUC-ROC metric on the validation dataset. ExtraTree algorithm had the highest AUC-ROC = 0.93 with sensitivity = 0.88 and precision = 0.92 in detecting ADRD status from audio-recorded data. Implications for current study: This study established the feasibility of utilizing spoken language to develop a screening algorithm with high performance for detecting patients with ADRD.

Figure 1. Analytic pipeline for modeling the spoken language



Aim 1. Model verbal communications of MCI-ED patients with nurses by developing an automated speech analysis system.

The analytic method of modeling patient-nurse verbal communication consists of five components.

Component 1: Modeling phonetic motor planning. Impairment in phonetic motor planning in patients with neurodegenerative disorders (e.g., MCI-ED) leads to poor pronunciation and alternation in phonological planning and speech rhythm.[43,44,45] We will use acoustic parameters in six domains (**Table 1**, **component 1**) to model the phonetic motor planning: i) alternation in speech fluency,[46,47,48] ii) rhythmic structure,[49–51] iii) frequency and spectral domain,[47,52,53] iv) voice instability,[47,54,55] v) voice quality,[56–58] and vi) voice intensity[47,59]

Component 2: Modeling the patient's emotional expression. Alternations in emotional expression developed in parallel with cognitive deterioration negatively influence the quality of the patient's interaction with others.[60,61] Emotion can be expressed in two major ways using nonverbal vocalization (vocally) and semantically via language. The underlying theoretical assumption for vocal expression of emotion is that affective processes (e.g., happiness or sadness) differentially change autonomic arousal and tension on vocal musculature, and these changes can be estimated by phonatory and articulatory parameters of the acoustic waveform.[62–64,65] To model the vocal expression of emotion, we use GeMAPS[38]. The GeMAPS parameters are presented in three domains (**Table 1, component 2**): i) frequency-related parameters, ii) energy/amplitude related parameters, and iii) spectral (balance/shape/dynamics) parameters. To model semantic expression of emotion, we will extract linguistic features indicating semantic expression of emotions using the LIWC (preliminary study D2.3).[40] The emotional expressions identified by LIWC (e.g., sadness, happiness) have been linked to healthcare outcomes (e.g., anxiety,[66] cognitive dysfunction[67]).

Table 1. Modeling MCI-ED patient-nurse verbal communication in the HHC setting (Components 1-4)

Component 1: Modeling phonetic motor planning	
Speech (Vocal) Fluency	i) Articulation is the number of phonemes per second without hesitation;[43] ii) Speech Rate is the number of phonemes per second with hesitation;[43] iii) Silent Pauses is the number of speechless interval at the beginning and between words;[47] iv) Within Word Disfluency is the within-word silent pauses and sound prolongations.[48]
Rhythmic Structure of Speech	i) Syllabic Intervals calculates temporal variability in speech;[68] ii) Pairwise Variability Index calculates durational variability in successive acoustic-phonetic intervals;[50] iv) Vowel Duration is the proportion of time of vocalic intervals in the sentence and the standard deviation of inter-vowel intervals.[51]
Frequency and Spectral Domain	i) Fundamental Frequency is the average of the number of oscillations that originates from the vocal fold per second;[69] ii) Formant Frequencies (F1, F2, F3, F4) are acoustic resonances of vocal tract that occurs due to the changes in the positions of vocal organs;[70] iii) Spectral Center of Gravity is the amplitude weighted mean of the harmonic peaks averaged over the sound duration;[71] iv)

	Long-Term Average Spectrum is a composite signal representing the spectrum of the glottal source as well as resonant characteristics of the vocal tract;[72] v) Mel-Frequency Cepstral Coefficients computes the energy variations between frequency bands of a speech signal.[73]
Voice Instability	i) Jitter is the measure of the cycle-to-cycle period variation of the successive glottal cycles;[74] ii) Shimmer is the cycle-to-cycle amplitude variations of the successive glottal cycles;[74] iii) Cepstral Peak Prominence counts for periodicity in the speech signal.[75]
Voice Quality	i) Harmonics-to-Noise Ratio quantifies the relative amount of additive noise in the voice signal;[69] ii) Voice Breaks measures the patient's reduced ability in vocal cord execution that can result in breaking voice;[76] iii) Acoustic Voice Quality Index consists of a weighted combination of time-frequency and quefrency-domain metrics that was originally developed to measure the severity of dysphonia.[77]
Voice Intensity	i) Hammarberg Index measures articulatory effort by computing the difference between the maximum energy in the 0…2kHz and the energy in the 2…5kHz band;[78,78] ii) Energy Concentration is the average of spectral frequency.[47]
Component 2: Modeling the patient's emotional expression	
Frequency Parameters	i) Pitch is the number of vibrations per second produced by the vocal cords;[79] ii) Jitter ;[74] iii) Centre Frequency of Formant 1, 2, and 3; and iv) Bandwidth of Formants 1, 2, and 3. Formant frequencies are acoustic resonances of the vocal tract that occur due to the changes in the positions of vocal organs.[70]
Energy/ Amplitude	i) Shimmer ; ii) Loudness is an estimate of perceived signal intensity from an auditory spectrum;[80] iii) Harmonics-to-Noise Ratio quantifies the relative amount of additive noise in the voice signal.[69]
Spectral Parameters	i) Alpha Ratio is the ratio of the summed energy from 50–1000 Hz and 1–5 kHz; ii) Hammarberg Index measures the articulatory effort;[78] iii) Spectral Slope 0–500 Hz and 500–1500 Hz; iv) Formant 1, 2, and 3 relative energy; v) Harmonic Difference H1-H2[81] is the difference between H1 (first harmonic amplitude) and H2 (second harmonic amplitude); vi) Harmonic Difference H1-A3 is the difference between H1 and A3 (energy of the highest harmonic in the third formant range);[81] vii) Spectral Flux is the difference of the spectra of two consecutive frames; viii) Mel-Frequency Cepstral Coefficients: see "Frequency and spectral domain" in component 1.
Component 3: Modeling syntactic, semantic, and pragmatic levels of language organization	
Lexical Richness	i) Moving Average Type-token Ratio[82] is the total number of unique words divided by the total number of words for each successive window of fixed length; ii) Brunet's Index is the variation in the type of words marked by part of speech tagging tool in a sentence with reference to the total number of words in a sentence;[83] iii) Honore Index measures the proportion of words that are used only once with reference to the total number of words.[84]

Syntactic Level of Language Organization	i) Level of Complexity of Sentences: we will compute the score of the complexity of patients sentences using a parse tree;[85] ii) Grammatical Errors in the Patient's Language: we will use a Parse Tree Analyzer to identify grammatical errors;[86] iii) Incomplete (fragment) Sentences: to compute this variable, we will use the Automatic Detection of Sentence Fragments Algorithm built on the combination of syntactic parse tree and part of speech tagging.[87]
Semantic Fluency	Semantic Fluency will be measured by identifying filled pauses (e.g., "um") in the patient's spoken language. [88,54,89,90]
Patient Recall Ability	i) Uncertainty in Patient Language will be computed using the linguistic approximator introduced by Ferson et al;[91] ii) Quantifying Memory Related Terms will be computed as a proportion of sentences with memory-related terms[92] with references to total number of sentences expressed by the patient using NimbleMiner toolkit[93] trained on a list of memory-related terms and their synonyms; iii) Question Ratio: we will use the NLTK package of Python to identify the proportion of interrogative sentences with reference to total number of sentences expressed by the patient.[92]
Component 4: Modeling patient-nurse interaction	
Patient Turns	Patient Turn is a continuous block of uninterrupted speech of a single patient. The total number of the patient's turns indicates the frequency of information exchange in the interaction.[94]
Interactivity	Dialog Interactivity is the total number of the patient's turns divided by the total length of the encounter.[94]
Turn Density	Turn Density will be computed using the same parameters specified for computing lexical richness (component 3).
Turn Duration	Turn Duration is the length of time of the patient's turn. It has been shown that MCI-ED patients have difficulty in monitoring their turns, and they may have a longer turn duration.[94]
The Relative Timing of Turns	i) Discernable Pause Rate is the proportion of discernible speechless intervals at the start of the patient's utterances (turns) with reference to total number of utterances;[95] ii) Cross-over Speaking Rate is the proportion of patient-nurse utterances with cross-over speaking with reference to the total number of utterances during the interaction.[95]

Component 3: Modeling syntactic, semantic, and pragmatic levels of language organization. Language impairment in MCI-ED patients is associated with aphasia-like symptoms and memory deficit symptoms, characterized by less dense and inaccurate speech planning,[96,97] difficulties in finding words,[43,98] simplified syntaxis and semantics,[43,67,99] and circumlocution.[43,100] These symptoms can result in communication errors and lower coherence in speech. We model language processing in MCI-ED patients using language parameters in four domains (**Table 1, component 3**): i) lexical richness,[83 84,101] ii) syntactic level of language organization,[85,102] iii) semantic fluency,[103] and iv) patient's recall ability.[91,92] We also use distilled RoBERTa[42] to represent the contextual relations among words or sub-words in the content of patient's spoken language.

Component 4: Modeling patient-nurse interaction. Patients with MCI-ED have difficulties identifying social cues required for smooth interaction.[104,105] The patient's communications have a recurrent pattern.[106] Therefore, analyzing naturally occurring patient-nurse interaction in the HHC setting enables us to identify the characteristic pattern of the patient's interaction that may indicate subtle cognitive decline in elderly patients. We use easily measured parameters of social interaction[94,95,107] to model the patient-nurse interactions in the HHC setting, including total number of patient turns, dialogue interactivity, turn density, turn duration, and relative timing of turns (**Table 1, component 4**).

Component 5: System implementation and evaluation. We compute acoustic parameters for Components 1 and 2 at the patient's utterance level (a continuous block of uninterrupted speech of a single patient) using OpenSMILE[36] and PRAAT Vocal Toolkits.[37]To compute linguistic parameters (Component 3), we will first use AWS Transcribe for automatic transcription of verbal communications. Then, we will compute linguistic parameters at the encounter level for each patient using the NLTK package and distilled RoBERTa. To compute parameters for modeling the patient-nurse interaction (Component 4), we will use the AWS Transcribe feature set, including timestamp and speaker identification.

Aim 2. Utilize large language models to automatically identify MCI-ED related information, including i) clinical symptoms, ii) lifestyle risk factors, and iii) communication deficits from both HHC clinical notes and patient-nurse verbal communication.

Component 1. Schema, ontology, and gold standard. We define a fixed information schema for three construct families—clinical symptoms, lifestyle risk factors, and communication deficits—capturing span text, canonical concept ID (SNOMED CT/UMLS where available), and attributes for assertion (present/absent/possible), temporality (current/historical), experiencer (patient/caregiver), severity, and frequency/duration. A seed taxonomy is compiled from clinical sources and expanded with corpus-driven variants (e.g., "losing words") to cover colloquial phrasing in notes and transcripts. A trained team of nurses and informatics RAs double-codes a stratified sample (by race, sex, visit type), targeting Cohen's $\kappa \geq 0.80$ for entity type and ≥ 0.75 for attributes before scaling. Adjudicated annotations form the gold standard, which we split into patient-disjoint train/dev/test sets (70/15/15) for modeling and evaluation.[108,109]

Component 2. Data preparation and de-identification. We de-identify PHI in HHC notes and normalize sections (e.g., History, Assessment, Plan), and we process encounter transcripts generated by AWS Transcribe by restoring punctuation, filtering noise tokens, and preserving speaker diarization for experiencer attribution. All texts undergo sentence splitting, tokenization, and linkage of medication, lab, and diagnosis mentions to RxNorm/LOINC/ICD-10 when they imply risk exposure.

Component 3. Modeling strategy. We use a hybrid approach with (A) prompted extraction and (B) lightweight adaptation.[108] For Track A, instruction-tuned LLMs are deployed in a HIPAA-compliant environment (Columbia AWS), using Llama-3.1-70B for highest-accuracy runs and Llama-3.1-8B for efficient iteration; structured prompts with a small number of curated in-domain exemplars and constrained decoding (function calling) enforce valid JSON outputs. We add self-consistency (multiple sampled extractions with voting) to reduce variance.[110] For Track B, we apply parameter-efficient fine-tuning (LoRA/QLoRA) on the gold standard to improve recall and attribute accuracy, preceded by brief continued pretraining on unlabeled HHC text to adapt to local style.

Component 4. Post-processing, normalization, and aggregation. We standardize each extracted mention to a SNOMED CT or UMLS concept using dictionary lookup with a similarity check, then resolve overlaps and duplicates by keeping the most specific asserted mention and discarding possible or negated ones.[108,111] Speaker tags and nearby context determine the experiencer (patient or caregiver); when ambiguous in narrative notes, we default to the patient. We aggregate results from encounters to visits and then to the patient level, setting patient-level flags (e.g., word-finding difficulty) when a construct appears in two or more encounters, and we retain dates for time-based analyses. The final output is a time-stamped patient table for integration with OASIS and Aim 1 speech variables.

Component 5. Evaluation and quality control. We report span-level precision, recall, and F1 under strict (exact) and relaxed (overlap) matching; macro-F1 for attributes (assertion, temporality, experiencer, severity); and concept-normalization accuracy. At the document and patient levels, we compute AUROC, AUPRC, and calibration (Brier score) for “any-evidence present” flags. All metrics are stratified by race, sex, and age bands; when disparities are observed, we apply mitigation such as re-weighting during PEFT and post-hoc threshold calibration, and we re-evaluate equality of opportunity/odds. Confidence scores from sequence log-probabilities and an auxiliary verifier flag low-confidence items for targeted human review, and weekly error-analysis updates prompts, PEFT settings, and normalization rules.

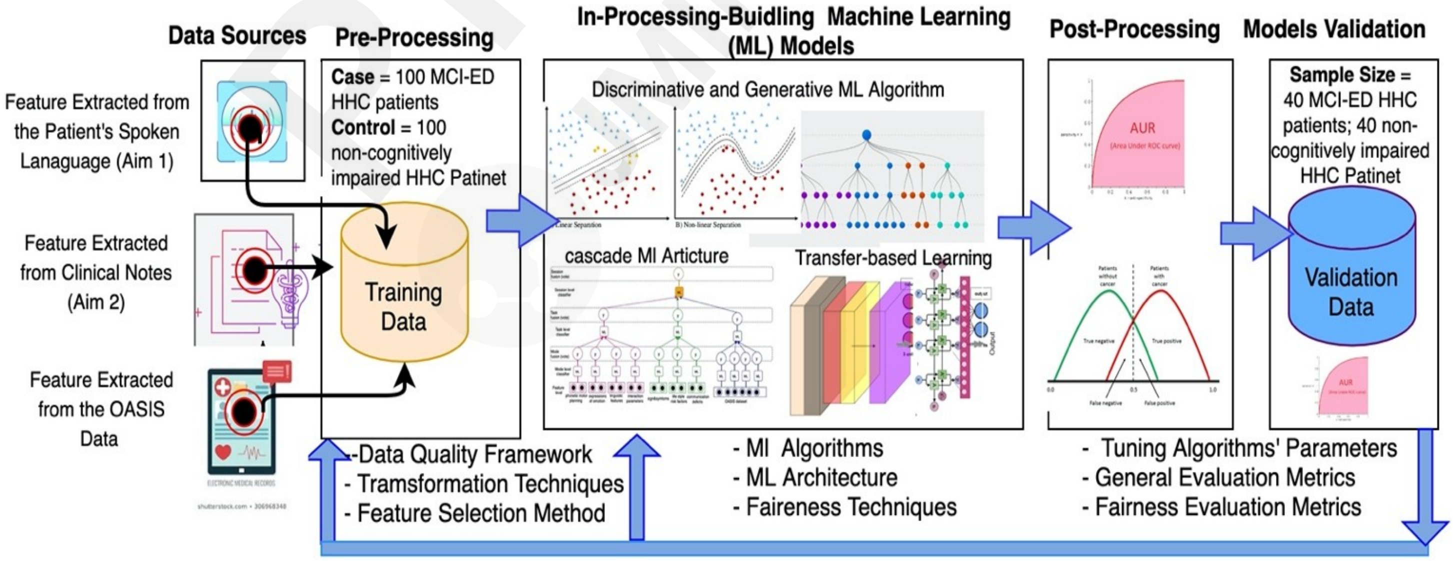
Component 6. Deployment, privacy, and reproducibility. Processing occurs on secure Columbia/VNS Health servers with role-based access, audit logging, and de-identified text for adaptation. We version prompts, PEFT checkpoints, and normalization resources; fix random seeds; and maintain a data card documenting corpus composition, preprocessing, and known limitations. Outputs and code are containerized to ensure repeatability across environments, and we prepare extract-transform-load scripts that merge LLM variables with OASIS and speech features for Aim 3.

Component 7. System output. After model selection on the development split and confirmation on the test split, we run the best configuration across all eligible HHC notes and encounter transcripts to generate per-patient summaries of clinical symptoms, lifestyle risk factors, and communication deficits with assertion/temporality, counts per visit, and normalized concept IDs. These patient-level tables are used for descriptive comparisons between cases and controls, correlation analyses with Aim 1 acoustic/interaction features, and as candidate predictors for the screening algorithm in Aim 3.

Aim 3: Develop a screening algorithm to identify HHC patients with MCI-ED.

The pipeline for developing this screening algorithm consists of four components (**Figure 2**), which will be built on the features extracted from patient-nurse verbal communication (Aim 1), clinical notes (Aim 2), and OASIS dataset.

Figure 2. Overview of the method for development of the MCI-ED screening algorithm



Data sources (input variables): Data sources for this study include all the acoustic and linguistic parameters extracted from the patient's spoken language (Aim 1), interaction parameters extracted from patient-nurse verbal interaction (Aim 1), clinical symptoms, lifestyle risk factors, communication deficits extracted from clinical notes and content of the patient-nurse verbal communication (Aim 2), and OASIS dataset including sociodemographic, diagnoses, medications, and functional status,

Component 1: Pre-Processing Phase. We will use the framework of data quality assessment that we developed at the Mayo Clinic for handling missing values, data inconsistency, and integrity issues.[112] Also, to boost the performance of ML algorithms, we will use feature transformation and scaling techniques such as log transform and Standard Scaler. As a feature selection method, we will use Joint Mutual Information Maximization (JMIM).[113] Compared to competing feature selection methods JMIM is has a higher generalization ability, particularly on small sample datasets with a large number of features.[113, 114]

Component 2: In-Processing Phase. (i) ML algorithms: To develop the MCI-ED screening algorithm, we will use (a) discriminative ML algorithms, including support vector machine,[115] multiple kernel learning,[116] generalized linear models[117] and ensembled decision trees (e.g., XGboost, Random Forest, and Extra-Trees). Discriminative ML algorithms in general are less prone to over-fitting and can effectively reduce the bias and variance-related errors of the screening algorithm.[117,118] Also, we will use (b) generative ML algorithms including Long Short-Term Memory Networks,[119] Hidden Markov model,[120] and Convolutional Neural Networks. The networks and model will be used to model the time dependency between dynamic (temporal) properties of speech parameters (Aim 1). Convolutional Neural Networks are effective models for modeling spectral correlations in acoustic parameters of speech.[121] (ii) ML architectures: We will use (a) cascade ML architecture that aims to build a multi-state ensemble learning based on concatenation of several ML algorithms in a cascade structure;[122] (b) multimodal ML architecture that aims to build an ML model that can process and join multiple data modalities (speech data, clinical notes, and OASIS) according to their importance in generating the probability of outcome class (presence of MCI-ED);[123] and (c) transfer-based learning[19] that focuses on transferring knowledge stored in an ML model from one domain to another. We will use the VGGish[124] and YAMNet[125] models, two pretrained acoustic detection models trained by Google on more than two million YouTube videos, to model dynamic properties in acoustic parameters of speech. To model the linguistic aspect of speech, we will use BERT[126] and its enhanced versions such as RoBERTa[42] and DistilBERT[127] to model the syntactic and semantic of language organization in the patient's spoken language.

Component 3: Post-Processing Phase. (i) To tune parameters of the ML algorithms and measure performance of the proposed ML architectures, we will use the stratified 5-fold nested cross validation method with a grid search algorithm. We will use AUC-ROC operating characteristics, Area Under Curve of Precision-recall (AUC-RP), and cumulative gain as general metrics for providing an overall picture of sensitivity, specificity, and precision of the models. We will also use fairness-based metrics,[128] including parity-based scores, equality of odds, treatment equality, and calibration score to measure outcome of different ML architectures on race groups and sex. (ii) If we observe that ML architectures have bias for a specific race or group, we will use different fairness approaches for mitigating bias.[129] These will include the reweighing approach,[130,131] adversarial learning,[132] and calibration.[133]

Component 4: Models Evaluation. In the final step, we will evaluate performance of the fine-tuned ML architectures on the validation dataset. We aim to achieve an AUC-ROC >0.85 for each race and sex group, which indicates a well-balanced and functioning screening algorithm. If the screening algorithm fails to achieve this performance, we will conduct another cycle/s of the system fine-tuning and explore performance of other ML algorithms and architectures until the desirable performance is achieved.

RESULTS

Initial Study Population and Clinical Characteristics

In this first phase of the proposed study, we enrolled 47 home healthcare patients with a balanced gender distribution (51.1% female) and diverse racial/ethnic representation (55.3% Black, 29.8% White, 8.5% Asian, 6.4% Hispanic). The majority were insured by Medicare (66.0%), and 44.7% lived alone. Early comparisons indicated that participants with symptoms of cognitive decline exhibited higher rates of urinary incontinence (36.8% vs. 21.4%), anxiety (57.9% vs. 39.3%), and impaired vision (26.3% vs. 7.1%), along with greater dependence in activities of daily living (73.7% completely dependent vs. 64.3%). Audio-recorded nurse–patient encounters averaged 19 minutes (IQR: 12–23) with a median of 56 utterances. Nurses consistently contributed more words than patients (median 842 vs. 589), underscoring their larger role in dialogue flow.

Preliminary Modeling Results

The initial modeling experiment demonstrated promising signals for distinguishing patients with cognitive decline. For patient speech alone, DistilBERT performed best among BERT variants (F1 = 69.39; AUC-ROC = 69.36), while BioClinicalBERT performed strongest on clinical notes (F1 = 64.29; AUC-ROC = 69.17). Among traditional classifiers, SVM-Linear achieved solid performance with patient speech (F1 = 75.0; AUC-ROC = 75.94). Adding nurse speech substantially improved results (SVM-F1 = 85.0; AUC-ROC = 86.47), suggesting that interactional features contribute meaningful information. Logistic Regression was most effective for EHR-based variables (F1 = 75.56; AUC-ROC = 79.70). When integrating all data streams—speech, interaction, and EHR—the SVM achieved the strongest overall results (F1 = 88.89; AUC-ROC = 90.23). Feature analysis highlighted reduced lexical diversity, longer patient silences, increased nurse dominance in speech, psycholinguistic markers, and EHR factors (e.g., non-insulin-dependent diabetes, pressure ulcers, living alone) as potential indicators of cognitive decline.

DISCUSSION

The proposed study is aligned with the NIA Strategic Directions for Research (2020-2025),[20] “Develop improved approaches for the early detection and diagnosis of disabling illnesses and age-related debilitating conditions” and addresses five of the directions recommended for future research on early detection of cognitive decline by an expert panel “Cost-Effective Early Detection of Cognitive Decline”[134] sponsored by the NIA: (1) identifying patients with cognitive impairment using routinely collected data, (2) clinical inference derived from speech and language processing, (3) using algorithms for disease detection, (4) designing low-cost cognitive assessment and screening tools for clinical settings, and (5) building non-invasive tests to detect symptoms of cognitive impairment. The recommendations of the panel emphasized the high rate of undiagnosed cognitive impairment and the strong potential of machine learning (ML) algorithms trained on multiple data streams, particularly routinely generated data in clinical settings, for developing a screening algorithm (screening tool) as a promising strategy to address this challenge.

Development of the MCI-ED screening algorithm is consistent with the goal of the Institute of Medicine for providing timely care for patients, and timely diagnosis is one of six aims for improving the quality of healthcare services.[135] In the HHC setting, nurses are uniquely positioned to provide timely care via tailored interventions for patients with cognitive impairment. However, our preliminary study showed that 63% of patients remained undiagnosed[136] due to inadequate transfer of information on patients’ cognitive status prior to entering HHC and inadequacy of the federally mandated, standard HHC assessment tool OASIS in identifying these patients.[25,137] The proposed screening algorithm has the potential to raise HHC clinicians’ attention to patients’ cognitive functioning for further evaluation and development of proper interventions such as addressing potential safety issues[12] to reduce negative outcomes.[138,139–141]

Although existing biomarkers, such as cerebrospinal fluid,[142] and magnetic resonance imaging[143] are sensitive enough for detecting cognitive impairment, their functionalities are limited due to high cost and perceived invasiveness.[14] Emerging studies show that a patient's spoken language is one of the earliest signs of cognitive decline, enabling spoken language to act as a biomarker to present multiple dimensions of cognitive abilities, including executive functioning, semantic memory, and language.[68,88,104] Patients with cognitive impairment have difficulty controlling vocal execution which effects the point, mode, and tension of articulation.[43] Cues of cognitive impairment conveyed in the voice have been empirically documented by measurement of the acoustic waveform (parameters), reflecting the shape of the vocal tract and the patient's abilities to control vocal cord execution during speech. The language of patients with cognitive impairment conveys cues that can be measured by linguistic parameters such as incomplete sentences[144] or filled pauses (e.g., "um").[145] A recent literature review on automatic detection of cognitive impairment from speech and language[146] showed that spoken language can provide promising results for detection of severe cognitive impairment (e.g., Alzheimer's), but findings for detecting MCI-ED were suboptimal,[46,147,47] due to limitations with study design and a small number of speech parameters. Most studies were conducted in a controlled setting with patients instructed to complete classical neuropsychological assessment tests (e.g., verbal fluency) in a short time (one minute).[100] Assessment tests are not sensitive enough to detect subtle cognitive changes in MCI-ED,[148] and their duration does not provide sufficient speech data to model all components of speech such as voice instability. We hypothesize that audio-recording patient-nurse's spontaneous communication during HHC episodes will provide sufficient data to model temporal parameters of speech components in MCI-ED patients, improving the screening algorithm's sensitivity.

MCI-ED is associated with a wide range of clinical symptoms[149] (e.g., low speech rate) and lifestyle risk factors (e.g., smoking).[150] Previous studies showed up to 80% of information related to patients' healthcare status is stored as free text (e.g., visit notes) using diverse vocabularies and that both explicit and implicit references can be captured using NLP methods.[151,152]Our preliminary studies demonstrate that large langue models (LLMs) holds promise for gaining insight into nurses' observations and patterns in clinical notes, but it has been underutilized for identifying MCI-ED related information (e.g., risk factors) within HHC clinical notes.[153,25,154] In addition to studying MCI-ED-related documentation patterns, our goal is to use LLM adaption strategies to explore the content of verbal communications of HHC patients-nurses for identifying potential uncaptured clinical risk factors in clinical notes and the OASIS dataset. We hypothesize that integrating speech parameters and MCI-ED related information extracted from verbal communications and clinical notes can improve the sensitivity of the screening algorithm.

Overall, this study featured five highly innovative components. First, the MCI-ED screening algorithm is unique as it will be built on multiple, easily accessible data streams generated in the HHC setting. Therefore, it has strong potential for integration into the HHC clinical workflow for timely identification of MCI-ED patients. Second, unlike previous studies conducted in controlled settings, we will take advantage of the MCI-ED patient's spontaneous speech during HHC episodes to identify temporal instability in speech components (e.g., speech rate) within and over encounters, which may occur due to the presence of neuropsychiatric symptoms, such as patient anxiety.[155] Modeling instability patterns in the acoustic, linguistic, and emotion components of speech have the potential to improve the sensitivity of the screening algorithm. Third, unlike previous studies,[68,46] we will model routine patient-nurse's spontaneous verbal interaction to extract informative interaction parameters that can be incorporated with other speech parameters to improve the sensitivity of the algorithm.[156] Fourth, this study is unique as it integrates multiple data streams, including MCI-ED related information extracted from clinical notes, speech parameters, and the OASIS dataset to address limitations of previous studies built on limited data sources.[144]

CONCLUSION

Developing an MCI-ED screening algorithm is a fundamental step for the timely identification of patients with cognitive impairment. The screening algorithm which uniquely takes advantage of easily accessible data generated in routine HHC encounters has the potential to be integrated into

HHC workflow to help HHC nurses in the early identification of patients with MCI-ED and referring them for further specialized neurological assessment. This work informs a future clinical decision support system for the early identification of MCI-ED in racially diverse HHC patients and evaluating its effectiveness on clinical outcome in clinical trials at VNS Health and HHC agencies in other regions. Additionally, we plan to create a public version of pipelines specified in the Aims for expanding its application for identification of other types of cognitive impairment (e.g., Alzheimer's disease) and identifying patients with communication difficulties.

REFERNCES

1. Zolnoori M, Zolnour A, Topaz M. ADscreen: A speech processing-based screening system for automatic identification of patients with Alzheimer's disease and related dementia. *Artif Intell Med*. 2023;143:102624.
2. Zolnoori M, Barrón Y, Song J, et al. HomeADScreen: Developing Alzheimer's disease and related dementia risk identification model in home healthcare. *Int J Med Inform*. Published online 2023:105146.
3. Zolnoori M, Zolnour A, Vergez S, et al. Beyond electronic health record data: leveraging natural language processing and machine learning to uncover cognitive insights from patient-nurse verbal communications. *Journal of the American Medical Informatics Association*. Published online 2024:ocae300.
4. Zhu CW, Sano M, Ferris SH, Whitehouse PJ, Patterson MB, Aisen PS. Health-related resource use and costs in elderly adults with and without mild cognitive impairment. *J Am Geriatr Soc*. 2013;61(3):396-402.
5. St-Hilaire A, Hudon C, Préville M, Potvin O. Utilization of healthcare services among elderly with cognitive impairment no dementia and influence of depression and anxiety: A longitudinal study. *Aging Ment Health*. 2017;21(8):810-822.
6. Zolnoori M, Sridharan S, Zolnour A, et al. Utilizing patient-nurse verbal communication in building risk identification models: the missing critical data stream in home healthcare. *Journal of the American Medical Informatics Association*. Published online 2023:ocad195.
7. Rovner BW, Casten RJ. Emergency department visits in African Americans with mild cognitive impairment and diabetes. *J Diabetes Complications*. 2021;35(5):107905.
8. Stephens CE, Newcomer R, Blegen M, Miller B, Harrington C. The effects of cognitive impairment on nursing home residents' emergency department visits and hospitalizations. *Alzheimer's & Dementia*. 2014;10(6):835-843.
9. Perry W, Lacritz L, Roebuck-Spencer T, et al. Population health solutions for assessing cognitive impairment in geriatric patients. *Innov Aging*. 2018;2(2):igy025.
10. Boise L, Neal MB, Kaye J. Dementia assessment in primary care: results from a study in three managed care systems. *J Gerontol A Biol Sci Med Sci*. 2004;59(6):M621-M626.
11. Tóth L, Hoffmann I, Gosztolya G, et al. A speech recognition-based solution for the automatic detection of mild cognitive impairment from spontaneous speech. *Curr Alzheimer Res*. 2018;15(2):130-138.
12. National Institute on Aging. Assessing Cognitive Impairment in Older Patients. Accessed March 1, 2021. https://www.nia.nih.gov/health/assessing-cognitive-impairment-older-patients
13. Lion KM, Szcześniak D, Bulińska K, et al. Do people with dementia and mild cognitive impairments experience stigma? A cross-cultural investigation between Italy, Poland and the UK. *Aging Ment Health*. 2020;24(6):947-955.
14. Zetterberg H, Blennow K. Blood biomarkers: Democratizing alzheimer's diagnostics. *Neuron*. 2020;106(6):881-883.

15. Judge D, Roberts J, Khandker R, Ambegaonkar B, Black CM. Physician perceptions about the barriers to prompt diagnosis of mild cognitive impairment and Alzheimer's disease. *Int J Alzheimers Dis*. 2019;2019.
16. Tarawneh R, Holtzman DM. The clinical problem of symptomatic Alzheimer disease and mild cognitive impairment. *Cold Spring Harb Perspect Med*. 2012;2(5):a006148.
17. Petersen RC. Mild cognitive impairment. *CONTINUUM: Lifelong Learning in Neurology*. 2016;22(2 Dementia):404.
18. Nation's Older Population to Nearly Double. Accessed March 7, 2021. https://www.census.gov/newsroom/press-releases/2014/cb14-84.html
19. Azadmaleki H, Haghbin Y, Rashidi S, et al. SpeechCARE: Harnessing Multimodal Innovation to Transform Cognitive Impairment Detection-Insights from the National Institute on Aging Alzheimer's Speech Challenge. *Stud Health Technol Inform*. 2025;329:1856-1857.
20. National Institute on Aging. The National Institute on Aging: Strategic Directions for Research, 2020-2025. 2021. https://www.nia.nih.gov/about/aging-strategic-directions-research/goal-health-interventions#c2
21. Ferris SH, Farlow M. Language impairment in Alzheimer's disease and benefits of acetylcholinesterase inhibitors. *Clin Interv Aging*. 2013;8:1007.
22. Jokel R, Lima BS, Fernandez A, Murphy KJ. Language in Amnestic Mild Cognitive Impairment and Dementia of Alzheimer's Type: Quantitatively or Qualitatively Different? *Dement Geriatr Cogn Dis Extra*. 2019;9(1):136-151.
23. Cera ML, Ortiz KZ, Bertolucci PHF, Minett T. Phonetic and phonological aspects of speech in Alzheimer's disease. *Aphasiology*. 2018;32(1):88-102.
24. Martin RC, Allen CM. A disorder of executive function and its role in language processing. In: *Seminars in Speech and Language*. Vol 29. NIH Public Access; 2008:201.
25. Topaz M, Adams V, Wilson P, Woo K, Ryvicker M. Free-Text Documentation of Dementia Symptoms in Home Healthcare: A Natural Language Processing Study. *Gerontol Geriatr Med*. 2020;6:2333721420959861.
26. Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc*. 2005;53(4):695-699.
27. Freitas S, Simões MR, Alves L, Santana I. Montreal cognitive assessment: validation study for mild cognitive impairment and Alzheimer disease. *Alzheimer Dis Assoc Disord*. 2013;27(1):37-43.
28. Charan J, Biswas T. How to calculate sample size for different study designs in medical research? *Indian J Psychol Med*. 2013;35(2):121.
29. Kinatukara S, Rosati RJ, Huang L. Assessment of OASIS reliability and validity using several methodological approaches. *Home health care services quarterly*. 2005;24(3):23-38.
30. Tullai-McGuinness S, Madigan EA, Fortinsky RH. Validity testing the Outcomes and Assessment Information Set (OASIS). *Home health care services quarterly*. 2009;28(1):45-57. doi:10.1080/01621420802716206
31. Zolnoori M, Vergez S, Kostic Z, et al. Audio recording patient-nurse verbal communications in home health care settings: pilot feasibility and usability study. *JMIR Hum Factors*. 2022;9(2):e35325.
32. 2.4GHz Wireless Lavalier Microphone,Saramonic Blink500 B6 Recording Mic Two Transmitters with battery-free Plug&Play USB Type-C Receiver for Samsung Galaxy, LG,HTC Google, and Other USB-C Type Devices. https://www.amazon.com/gp/product/B081Z17W38/ref=ox_sc_act_title_22?smid=AU14N51PPEZY0&psc=1
33. Zolnoori M, Vergez S, Sridharan S, et al. Is the patient speaking or the nurse? Automatic speaker type identification in patient-nurse audio recordings. *J Am Med Inform Assoc*. 2023;30(10):1673-1683. doi:10.1093/JAMIA/OCAD139
34. RX 8-Great Audio Starts With RX. https://www.izotope.com/en/products/rx.html
35. Zolnoori M, Zolnour A, Topaz M. ADscreen: A speech processing-based screening system for

automatic identification of patients with Alzheimer's disease and related dementia. *Artif Intell Med*. 2023;143:102624.
36. Eyben F, Schuller B. openSMILE:) The Munich open-source large-scale multimedia feature extractor. *ACM SIGMultimedia Records*. 2015;6(4):4-13.
37. Praat Vocal Toolkit. Accessed April 9, 2021. http://www.praatvocaltoolkit.com/
38. Eyben F, Scherer KR, Schuller BW, et al. The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. *IEEE Trans Affect Comput*. 2015;7(2):190-202.
39. Shahin M, Ahmed B, Smith D V, Duenser A, Epps J. Automatic Screening Of Children With Speech Sound Disorders Using Paralinguistic Features. In: *2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE; 2019:1-5.
40. Pennebaker JW, Boyd RL, Jordan K, Blackburn K. *The Development and Psychometric Properties of LIWC2015*.; 2015.
41. Bird S, Klein E, Loper E. *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. " O'Reilly Media, Inc."; 2009.
42. Liu Y, Ott M, Goyal N, et al. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:190711692*. Published online 2019.
43. Meilán JJG, Martínez-Sánchez F, Martínez-Nicolás I, Llorente TE, Carro J. Changes in the rhythm of speech difference between people with nondegenerative mild cognitive impairment and with preclinical dementia. *Behavioural neurology*. 2020;2020.
44. Duffy JR, Josephs KA. The diagnosis and understanding of apraxia of speech: Why including neurodegenerative etiologies may be important. *Journal of Speech, Language, and Hearing Research*. Published online 2012.
45. Ward M, Cecato JF, Aprahamian I, Martinelli JE. Assessment for apraxia in Mild Cognitive Impairment and Alzheimer's disease. *Dement Neuropsychol*. 2015;9(1):71-75.
46. Nagumo R, Zhang Y, Ogawa Y, et al. Automatic detection of cognitive impairments through acoustic analysis of speech. *Curr Alzheimer Res*. 2020;17(1):60.
47. Themistocleous C, Eckerström M, Kokkinakis D. Voice quality and speech fluency distinguish individuals with Mild Cognitive Impairment from Healthy Controls. *PLoS One*. 2020;15(7):e0236009.
48. Huet K, Delvaux V, Piccaluga M, Roland V, Harmegnies B. Inter-Syllabic Interval as an indicator of fluency in Parkinsonian French speech. In: *11th International Seminar on Speech Production, Tianjin, China*. ; 2017.
49. Beltrami D, Gagliardi G, Rossini Favretti R, Ghidoni E, Tamburini F, Calzà L. Speech Analysis by Natural Language Processing Techniques: A Possible Tool for Very Early Detection of Cognitive Decline? *Frontiers in Aging Neuroscience*. 2018;10:369. doi:10.3389/fnagi.2018.00369
50. Calzà L, Gagliardi G, Favretti RR, Tamburini F. Linguistic features and automatic classifiers for identifying mild cognitive impairment and dementia. *Computer Speech & Language*. 2021;65:101113.
51. Dahmani H, Selouani SA, O'shaughnessy D, Chetouani M, Doghmane N. Assessment of dysarthric speech through rhythm metrics. *Journal of King Saud University-Computer and Information Sciences*. 2013;25(1):43-49.
52. 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.
53. Bidelman GM, Lowther JE, Tak SH, Alain C. Mild cognitive impairment is characterized by deficient brainstem and cortical representations of speech. *Journal of Neuroscience*. 2017;37(13):3610-3620.
54. López-de-Ipiña K, Martinez-de-Lizarduy U, Calvo PM, et al. On the analysis of speech and disfluencies for automatic detection of Mild Cognitive Impairment. *Neural Comput Appl*. Published online 2018:1-9.

55. Yu B, Williamson JR, Mundt JC, Quatieri TF. Speech-based automated cognitive impairment detection from remotely-collected cognitive test audio. *IEEE Access*. 2018;6:40494-40505.
56. Al-Hameed S, Benaissa M, Christensen H, Mirheidari B, Blackburn D, Reuber M. A new diagnostic approach for the identification of patients with neurodegenerative cognitive complaints. *PloS one*. 2019;14(5):e0217388.
57. Mirzaei S, El Yacoubi M, Garcia-Salicetti S, et al. Two-Stage Feature Selection of Voice Parameters for Early Alzheimer's Disease Prediction. *IRBM*. 2018;39(6):430-435. doi:10.1016/j.irbm.2018.10.016
58. König A, Satt A, Sorin 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):112-124.
59. Ning L, Luo K. Using text and acoustic features to diagnose Mild Cognitive Impairment and Alzheimer's disease. Published online 2020.
60. Han KH, Zaytseva Y, Bao Y, et al. Impairment of vocal expression of negative emotions in patients with Alzheimer's disease. *Front Aging Neurosci*. 2014;6:101.
61. Cadieux NL, Greve KW. Emotion processing in Alzheimer's disease. *Journal of the International Neuropsychological Society*. 1997;3(5):411-419.
62. Kamiloğlu RG, Fischer AH, Sauter DA. Good vibrations: A review of vocal expressions of positive emotions. *Psychon Bull Rev*. 2020;27(2):237-265.
63. 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.
64. Spazzapan EA, Marino VC de C, Cardoso VM, Berti LC, Fabron EMG. Acoustic characteristics of voice in different cycles of life: an integrative literature review. *Revista CEFAC*. 2019;21(3).
65. Grabowski K, Rynkiewicz A, Lassalle A, et al. Emotional expression in psychiatric conditions: New technology for clinicians. *Psychiatry and clinical neurosciences*. 2019;73(2):50-62.
66. Shen JH, Rudzicz F. Detecting anxiety through reddit. In: *Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology—From Linguistic Signal to Clinical Reality*. ; 2017:58-65.
67. Asgari M, Kaye J, Dodge H. Predicting mild cognitive impairment from spontaneous spoken utterances. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*. 2017;3(2):219-228.
68. Martínez-Nicolás I, Llorente TE, Martínez-Sánchez F, Meilán JJG. Ten Years of Research on Automatic Voice and Speech Analysis of People With Alzheimer's Disease and Mild Cognitive Impairment: A Systematic Review Article. *Front Psychol*. 2021;12:645.
69. Boersma P. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In: *Proceedings of the Institute of Phonetic Sciences*. Vol 17. Amsterdam; 1993:97-110.
70. Viegas F, Viegas D, Guimarães GS, et al. Comparison of fundamental frequency and formants frequency measurements in two speech tasks. *Revista CEFAC*. 2019;21(6).
71. Kong YY, Mullangi A, Marozeau J, Epstein M. Temporal and spectral cues for musical timbre perception in electric hearing. Published online 2011.
72. Tjaden K, Sussman JE, Liu G, Wilding G. Long-term average spectral (LTAS) measures of dysarthria and their relationship to perceived severity. *Journal of Medical Speech-Language Pathology*. 2010;18(4):125.
73. On CK, Pandiyan PM, Yaacob S, Saudi A. Mel-frequency cepstral coefficient analysis in speech recognition. In: *2006 International Conference on Computing & Informatics*. IEEE; 2006:1-5.
74. Teixeira JP, Oliveira C, Lopes C. Vocal acoustic analysis–jitter, shimmer and hnr parameters. *Procedia Technology*. 2013;9:1112-1122.
75. Fraile R, Godino-Llorente JI. Cepstral peak prominence: A comprehensive analysis. *Biomedical Signal Processing and Control*. 2014;14:42-54.

76. Simonyan K, Tovar-Moll F, Ostuni J, et al. Focal white matter changes in spasmodic dysphonia: a combined diffusion tensor imaging and neuropathological study. *Brain*. 2008;131(2):447-459.
77. Maryn Y, De Bodt M, Roy N. The Acoustic Voice Quality Index: toward improved treatment outcomes assessment in voice disorders. *Journal of communication disorders*. 2010;43(3):161-174.
78. Tamarit L, Goudbeek M, Scherer K. Spectral slope measurements in emotionally expressive speech. *Proceedings of Speech Analysis and Processing for Knowledge Discovery*. Published online 2008:169-183.
79. Deutsch D, Henthorn T, Dolson M. Absolute pitch, speech, and tone language: Some experiments and a proposed framework. *Music perception*. 2004;21(3):339-356.
80. Gramming P, Sundberg J, Ternström S, Leanderson R, Perkins WH. Relationship between changes in voice pitch and loudness. *Journal of voice*. 1988;2(2):118-126.
81. Narasimhan S V, Vishal K. Spectral measures of hoarseness in persons with hyperfunctional voice disorder. *Journal of Voice*. 2017;31(1):57-61.
82. Fergadiotis G, Wright HH, Green SB. Psychometric evaluation of lexical diversity indices: Assessing length effects. *Journal of Speech, Language, and Hearing Research*. 2015;58(3):840-852.
83. Sanborn V, Ostrand R, Ciesla J, Gunstad J. Automated assessment of speech production and prediction of MCI in older adults. *Appl Neuropsychol Adult*. Published online 2020:1-8.
84. Ntracha A, Iakovakis D, Hadjidimitriou S, Charisis VS, Tsolaki M, Hadjileontiadis LJ. Detection of Mild Cognitive Impairment Through Natural Language and Touchscreen Typing Processing. Front. Digit. *Health N Hav*. 2020;2:567158.
85. Cheung H, Kemper S. Competing complexity metrics and adults' production of complex sentences. *Applied Psycholinguistics*. 1992;13(01):53e76.
86. Lakretz Y, Dehaene S, King JR. What Limits Our Capacity to Process Nested Long-Range Dependencies in Sentence Comprehension? *Entropy*. 2020;22(4):446.
87. Yeung CY, Lee J. Automatic detection of sentence fragments. In: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. ; 2015:599-603.
88. Tóth L, Gosztolya G, Vincze V, et al. Automatic detection of mild cognitive impairment from spontaneous speech using ASR. In: *Sixteenth Annual Conference of the International Speech Communication Association*. ; 2015.
89. Xu Z, Vergez S, Esmaeili E, et al. Voice for All: Evaluating the Accuracy and Equity of Automatic Speech Recognition Systems in Transcribing Patient Communications in Home Healthcare. *Stud Health Technol Inform*. 2025;329:1904-1906.
90. Zolnoori M, Vergez S, Xu Z, et al. Decoding disparities: evaluating automatic speech recognition system performance in transcribing Black and White patient verbal communication with nurses in home healthcare. *JAMIA Open*. 2024;7(4):ooae130.
91. Ferson S, O'Rawe J, Antonenko A, et al. Natural language of uncertainty: numeric hedge words. *International Journal of Approximate Reasoning*. 2015;57:19-39.
92. Khodabakhsh A, Yesil F, Guner E, Demiroglu C. Evaluation of linguistic and prosodic features for detection of Alzheimer's disease in Turkish conversational speech. *EURASIP Journal on Audio, Speech, and Music Processing*. 2015;2015(1):9.
93. Topaz M. NimbleMiner: A Novel Multi-Lingual Text Mining Application. *Studies in health technology and informatics*. 2019;264:1608-1609.
94. Roter DL, Larson SM, Beach MC, Cooper LA. Interactive and evaluative correlates of dialogue sequence: a simulation study applying the RIAS to turn taking structures. *Patient Educ Couns*. 2008;71(1):26-33.
95. Drew P, Chatwin J, Collins S. Conversation analysis: a method for research into interactions between patients and health-care professionals. *Health Expectations*. 2001;4(1):58-70.
96. Roark B, Mitchell M, Hosom JP, Hollingshead K, Kaye J. Spoken language derived measures for detecting mild cognitive impairment. *IEEE Trans Audio Speech Lang Process*. 2011;19(7):2081-

2090.
97. Kim BS, Kim YB, Kim H. Discourse measures to differentiate between mild cognitive impairment and healthy aging. *Front Aging Neurosci*. 2019;11:221.
98. Aramaki E, Shikata S, Miyabe M, Kinoshita A. Vocabulary size in speech may be an early indicator of cognitive impairment. *PLoS One*. 2016;11(5):e0155195.
99. Sung JE, Choi S, Eom B, Yoo JK, Jeong JH. Syntactic complexity as a linguistic marker to differentiate mild cognitive impairment from normal aging. *Journal of Speech, Language, and Hearing Research*. 2020;63(5):1416-1429.
100. Mueller KD, Hermann B, Mecollari J, Turkstra LS. Connected speech and language in mild cognitive impairment and Alzheimer's disease: A review of picture description tasks. *J Clin Exp Neuropsychol*. 2018;40(9):917-939.
101. Ostrand R, Gunstad J. Using automatic assessment of speech production to predict current and future cognitive function in older adults. *Journal of Geriatric Psychiatry and Neurology*. Published online 2020:0891988720933358.
102. Davy W, Travis JA, Laura W, Dona L, Graciela G. Towards automatic detection of abnormal cognitive decline and dementia through linguistic analysis of writing samples. In: *Proceedings of the 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics*. ; 2016:1198-1207.
103. Gosztolya G, Tóth L, Grósz T, et al. Detecting mild cognitive impairment from spontaneous speech by correlation-based phonetic feature selection. Published online 2016.
104. Johnson M, Lin F. Communication difficulty and relevant interventions in mild cognitive impairment: implications for neuroplasticity. *Top Geriatr Rehabil*. 2014;30(1):18.
105. Pan CW, Wang X, Ma Q, Sun HP, Xu Y, Wang P. Cognitive dysfunction and health-related quality of life among older Chinese. *Sci Rep*. 2015;5(1):1-8.
106. Ilan S, Carmel S. Patient communication pattern scale: psychometric characteristics. *Health Expectations*. 2016;19(4):842-853.
107. H Dodge H, Mattek N, Gregor M, et al. Social markers of mild cognitive impairment: Proportion of word counts in free conversational speech. *Curr Alzheimer Res*. 2015;12(6):513-519.
108. Zhang Z, Gupta P, Song J, Zolnoori M, Topaz M. From Conversation to Standardized Terminology: An LLM-RAG Approach for Automated Health Problem Identification in Home Healthcare. *J Nurs Scholarsh*. Published online 2025.
109. Zhang Z, Nezhad MJM, Gupta P, Topaz M, Zolnoori M. Performance of LLMs in Citation Screening: A Comparison Across Datasets with Varied Inclusion Rates. *Stud Health Technol Inform*. 2025;329:1886-1887.
110. Zolnour A, Azadmaleki H, Haghbin Y, et al. LLMCARE: Alzheimer's Detection via Transformer Models Enhanced by LLM-Generated Synthetic Data. *arXiv preprint arXiv:250810027*. Published online 2025.
111. Hosseini SMB, Nezhad MJM, Hosseini M, Zolnoori M. Optimizing Entity Recognition in Psychiatric Treatment Data with Large Language Models. *Stud Health Technol Inform*. 2025;329:784-788.
112. Zolnoori M, Williams MD, Leasure WB, Angstman KB, Ngufor C. A Systematic Framework for Analyzing Observation Data in Patient-Centered Registries: Case Study for Patients With Depression. *JMIR Res Protoc*. 2020;9(10):e18366.
113. Bennasar M, Hicks Y, Setchi R. Feature selection using Joint Mutual Information Maximisation. *Expert Syst Appl*. 2015;42(22):8520-8532. doi:10.1016/j.eswa.2015.07.007
114. Varatharajah Y, Ramanan VK, Iyer R, Vemuri P. Predicting short-term MCI-to-AD progression using imaging, CSF, genetic factors, cognitive resilience, and demographics. *Sci Rep*. 2019;9(1):1-15.
115. Pradhan A. Support vector machine-a survey. *International Journal of Emerging Technology and Advanced Engineering*. 2012;2(8):82-85.

116. Castro E, Gómez-Verdejo V, Martínez-Ramón M, Kiehl KA, Calhoun VD. A multiple kernel learning approach to perform classification of groups from complex-valued fMRI data analysis: application to schizophrenia. *Neuroimage*. 2014;87:1-17.
117. Engebretsen S, Bohlin J. Statistical predictions with glmnet. *Clin Epigenetics*. 2019;11(1):1-3.
118. Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*. ; 2016:785-794.
119. Sherstinsky A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D*. 2020;404:132306.
120. Eddy SR. What is a hidden Markov model? *Nat Biotechnol*. 2004;22(10):1315-1316.
121. Zhang Y, Pezeshki M, Brakel P, Zhang S, Bengio CLY, Courville A. Towards end-to-end speech recognition with deep convolutional neural networks. *arXiv preprint arXiv:170102720*. Published online 2017.
122. Pakrashi A, Mac Namee B. CascadeML: An Automatic Neural Network Architecture Evolution and Training Algorithm for Multi-label Classification (Best Technical Paper). In: *International Conference on Innovative Techniques and Applications of Artificial Intelligence*. Springer; 2019:3-17.
123. Baltrušaitis T, Ahuja C, Morency LP. Multimodal machine learning: A survey and taxonomy. *IEEE Trans Pattern Anal Mach Intell*. 2018;41(2):423-443.
124. Hershey S, Chaudhuri S, Ellis DPW, et al. CNN architectures for large-scale audio classification. In: *2017 Ieee International Conference on Acoustics, Speech and Signal Processing (Icassp)*. IEEE; 2017:131-135.
125. Gemmeke JF, Ellis DPW, Freedman D, et al. Audio set: An ontology and human-labeled dataset for audio events. In: *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE; 2017:776-780.
126. Devlin J, Chang MW, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:181004805*. Published online 2018.
127. Sanh V, Debut L, Chaumond J, Wolf T. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:191001108*. Published online 2019.
128. Garg P, Villasenor J, Foggo V. Fairness metrics: A comparative analysis. In: *2020 IEEE International Conference on Big Data (Big Data)*. IEEE; 2020:3662-3666.
129. Rashidi S, Azadmaleki H, Zolnour A, Nezhad MJM, Zolnoori M. SpeechCura: A Novel Speech Augmentation Framework to Tackle Data Scarcity in Healthcare. *Stud Health Technol Inform*. 2025;329:1858-1859.
130. Krasanakis E, Spyromitros-Xioufis E, Papadopoulos S, Kompatsiaris Y. Adaptive sensitive reweighting to mitigate bias in fairness-aware classification. In: *Proceedings of the 2018 World Wide Web Conference*. ; 2018:853-862.
131. Jiang H, Nachum O. Identifying and correcting label bias in machine learning. In: *International Conference on Artificial Intelligence and Statistics*. PMLR; 2020:702-712.
132. Wadsworth C, Vera F, Piech C. Achieving fairness through adversarial learning: an application to recidivism prediction. *arXiv preprint arXiv:180700199*. Published online 2018.
133. Pleiss G, Raghavan M, Wu F, Kleinberg J, Weinberger KQ. On fairness and calibration. *arXiv preprint arXiv:170902012*. Published online 2017.
134. Bhattacharyya P, W Bynum JP, Carrillo M, et al. *Cost-Effective Early Detection of Cognitive Decline*.; 2017.
135. Bradford A, Kunik ME, Schulz P, Williams SP, Singh H. Missed and delayed diagnosis of dementia in primary care: prevalence and contributing factors. *Alzheimer Dis Assoc Disord*. 2009;23(4):306.
136. Ryvicker M, Barrón Y, Song J, et al. Using Natural Language Processing to Identify Home Health Care Patients at Risk for Diagnosis of Alzheimer's Disease and Related Dementias. *Journal of Applied Gerontology*. Published online 2024:07334648241242321.
137. Ryvicker M, Barrón Y, Shah S, et al. Clinical and Demographic Profiles of Home Care Patients With

Alzheimer’s Disease and Related Dementias: Implications for Information Transfer Across Care Settings. *Journal of Applied Gerontology*. Published online 2021:0733464821999225.
138. Ryvicker M, Barrón Y, Shah S, et al. Clinical and Demographic Profiles of Home Care Patients With Alzheimer’s Disease and Related Dementias: Implications for Information Transfer Across Care Settings. *Journal of Applied Gerontology*. Published online 2021:0733464821999225.
139. Bickel H, Hendlmeier I, Heßler JB, et al. The Prevalence of Dementia and Cognitive Impairment in Hospitals: Results from the General Hospital Study (GHoSt). *Dtsch Arztebl Int*. 2018;115(44):733.
140. Launay C, Haubois G, Hureaux-Huynh R, Gautier J, Annweiler C, Beauchet O. Older adults and emergency department: who is at risk of hospitalization? *Geriatr Psychol Neuropsychiatr Vieil*. 2014;12(1):43-49.
141. Gray LC, Peel NM, Costa AP, et al. Profiles of older patients in the emergency department: findings from the interRAI Multinational Emergency Department Study. *Ann Emerg Med*. 2013;62(5):467-474.
142. Attier-Zmudka J, Sérot JM, Valluy J, et al. Decreased cerebrospinal fluid flow is associated with cognitive deficit in elderly patients. *Front Aging Neurosci*. 2019;11:87.
143. Van De Pol LA, Korf ESC, Van Der Flier WM, et al. Magnetic resonance imaging predictors of cognition in mild cognitive impairment. *Arch Neurol*. 2007;64(7):1023-1028.
144. Fraser KC, Lundholm Fors K, Eckerström M, Öhman F, Kokkinakis D. Predicting MCI status from multimodal language data using cascaded classifiers. *Front Aging Neurosci*. 2019;11:205.
145. Mirzaei S, El Yacoubi M, Garcia-Salicetti S, et al. Two-stage feature selection of voice parameters for early Alzheimer’s disease prediction. *IRBM*. 2018;39(6):430-435.
146. Petti U, Baker S, Korhonen A. A systematic literature review of automatic Alzheimer’s disease detection from speech and language. *Journal of the American Medical Informatics Association*. 2020;27(11):1784-1797.
147. Beltrami D, Calzà L, Gagliardi G, et al. Automatic identification of mild cognitive impairment through the analysis of Italian spontaneous speech productions. In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*. ; 2016:2086-2093.
148. Howieson D. Current limitations of neuropsychological tests and assessment procedures. *Clin Neuropsychol*. 2019;33(2):200-208.
149. Langa KM, Levine DA. The diagnosis and management of mild cognitive impairment: a clinical review. *Jama*. 2014;312(23):2551-2561.
150. Huan-qing YU, Zong-junA GUO, Xiao-lin W, et al. Exploration of high risk factors in lifestyle for mild cognitive impairment in elderly people. *Progress in Modern Biomedicine*. 2011;11(10):1885-1890.
151. Ford E, Carroll JA, Smith HE, Scott D, Cassell JA. Extracting information from the text of electronic medical records to improve case detection: a systematic review. *Journal of the American Medical Informatics Association*. 2016;23(5):1007-1015.
152. Meystre SM, Savova GK, Kipper-Schuler KC, Hurdle JF. Extracting information from textual documents in the electronic health record: a review of recent research. *Yearb Med Inform*. 2008;17(01):128-144.
153. Navathe AS, Zhong F, Lei VJ, et al. Hospital Readmission and Social Risk Factors Identified from Physician Notes. *Health Services Research*. 2018;53(2):1110-1136. doi:10.1111/1475-6773.12670
154. Zhang Z, Nezhad MJM, Hosseini SMB, et al. A Scoping Review of Large Language Model Applications in Healthcare. *Stud Health Technol Inform*. 2025;329:1966-1967.
155. Cerejeira J, Lagarto L, Mukaetova-Ladinska E. Behavioral and psychological symptoms of dementia. *Front Neurol*. 2012;3:73.
156. Meilán JJG, Martínez-Sánchez F, Carro J, López DE, Millian-Morell L, Arana JM. Speech in Alzheimer’s disease: Can temporal and acoustic parameters discriminate dementia? *Dement Geriatr Cogn Disord*. 2014;37(5-6):327-334.

Supplementary Files

Manuscript - Proposal.
URL: http://asset.jmir.pub/assets/4049151c42bbb1194c4a192c48d40719.pdf

Multimedia Appendixes

Reviewers' comments for the proposal.
URL: http://asset.jmir.pub/assets/609d3d029302380150baf1ff7f46b84d.pdf