Protocol

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

Maryam Zolnoori, PhD

Columbia University Irving Medical Center, New York, NY, United States

Corresponding Author:
Maryam Zolnoori, PhD
Columbia University Irving Medical Center
560 W 168th St
New York, NY, 10032
United States
Phone: 1 317 515 1950
Email: m.zolnoori@gmail.com

Abstract

Background: Mild cognitive impairment and early dementia (MCI-ED) are frequently unrecognized in routine care, particularly in home health care (HHC), where clinical decisions are made under time constraints and cognitive status may be incompletely documented. Federally mandated HHC assessments, such as the Outcome and Assessment Information Set (OASIS), capture health and functional status but may miss subtle early cognitive changes. Speech, language, and interactional patterns during routine patient-nurse communication, together with information embedded in unstructured clinical notes, may provide complementary signals for earlier identification.

Objective: This protocol describes the development and evaluation of a multimodal screening approach for identifying MCI-ED in HHC by integrating (1) speech and interaction features from routine patient-nurse encounters (verbal communication), (2) large language model–based extraction of MCI-ED–related information from HHC notes and encounter transcripts, and (3) structured variables from OASIS.

Methods: This ongoing cross-sectional case-control study is being conducted in collaboration with VNS Health (formerly Visiting Nurse Service of New York). Eligible participants are adults aged ≥60 years receiving HHC services. Case/control assignment uses a 2-stage process: electronic health record (EHR) prescreening followed by clinician-reviewed cognitive assessment (Montreal Cognitive Assessment and Clinical Dementia Rating) for consented participants without an existing mild cognitive impairment diagnosis. For Aim 1, each participant contributes 3 audio-recorded routine patient-nurse encounters linked to EHR data, including OASIS and free-text clinical notes. Aim 1 extracts acoustic, linguistic, emotional, and interactional features from patient-nurse verbal communication. Aim 2 uses a schema-guided large language model pipeline to extract and normalize MCI-ED–related symptoms, lifestyle risk factors, and communication deficits from HHC notes and encounter transcripts, supported by a human-annotated gold-standard dataset. Aim 3 integrates speech, extracted text variables, and OASIS predictors using supervised machine learning with stratified nested cross-validation; evaluation will include discrimination, calibration, and subgroup performance checks across race, sex, and age.

Results: Between February 2024 and July 2025, a total of 114 HHC patients completed study-administered cognitive assessments and were classified as 55 MCI-ED cases and 59 cognitively normal controls. Audio-recorded patient-nurse encounters had a median duration of 19 (IQR 12-23) minutes and a median of 56 (IQR 31-80) utterances per encounter; nurses contributed more words than patients (median 842, IQR 461-1218 vs median 589, IQR 303-960). In exploratory feasibility analyses, multimodal models integrating speech, interactional features, and structured EHR/OASIS variables outperformed single-source models.

Conclusions: This protocol describes a reproducible multimodal framework for MCI-ED screening in HHC using routinely generated data streams. Initial implementation results support feasibility of data collection and end-to-end processing and suggest potential value of integrating interactional speech features with clinical text and OASIS variables. Final model evaluation, subgroup analyses, and validation will follow the prespecified analytic procedures on the finalized study dataset.

International Registered Report Identifier (IRRID): DERR1-10.2196/82731

Introduction

Alzheimer disease and Alzheimer disease–related dementias are among the most pressing global public health challenges. In 2021, an estimated 57 million people were living with dementia worldwide, with over 60% residing in low- and middle-income countries and nearly 10 million new cases occurring each year [1]. Dementia is associated with substantial disability, caregiver burden, and rapidly rising health-system costs [2]. In the United States, an estimated 7.2 million adults aged ≥60 years were living with Alzheimer dementia in 2025, underscoring the scale of need in high-income settings as well [3]. Alongside treatment advances and growing interest in risk reduction, global guidance continues to emphasize the importance of timely detection and prevention-oriented interventions across diverse populations and care contexts [4,5].

Despite this urgency, a large proportion of cognitive impairment, particularly mild cognitive impairment (MCI) and early-stage dementia, remains unrecognized or undocumented in routine care, contributing to delayed diagnosis and missed opportunities to tailor care planning [6,7]. In home-based care settings, documentation gaps can be especially consequential because care teams must make clinical decisions in the context of multimorbidity, limited visit time, and incomplete prior cognitive history [8,9]. Recent evidence from skilled home health care (HHC) shows that dementia is frequently undocumented in home health records, illustrating how care transitions and documentation practices can impede recognition of cognitive impairment [10]. These realities motivate scalable screening approaches that can operate within routine workflows, rather than relying solely on specialist evaluation or resource-intensive testing.

Speech and language have emerged as promising noninvasive, low-burden digital biomarkers for cognitive impairment [11-14]. A recent systematic review and meta-analysis focused on MCI specifically concluded that speech-based biomarkers show meaningful diagnostic use, while also highlighting methodological heterogeneity and the need for validation in diverse settings and populations [15]. At the same time, much of the speech-based Alzheimer disease and Alzheimer disease–related dementias detection literature remains anchored in structured elicitation tasks [16] (eg, cookie-theft picture description) and benchmark corpora (eg, DementiaBank-derived shared tasks such as ADReSSo [17]), which enable comparability but may not capture interactional and pragmatic markers expressed during everyday clinical communication [15,18]. Multilingual and cross-cultural work further indicates that generalization across languages and contexts cannot be assumed; for example, multilingual spontaneous speech studies (eg, Italian and Spanish) demonstrate feasibility outside English-centric benchmarks but also reinforce the importance of ecologically valid sampling and external validation [13,19-21].

These limitations are particularly relevant for MCI and early dementia, where impairments can be subtle, context-dependent, and potentially expressed through conversational dynamics (eg, timing, turn-taking balance, and discourse coherence) rather than only through content produced during structured tasks [16]. This motivates studying routine patient-clinician conversations, where interactional features may provide an additional signal for early-stage cognitive change in real-world contexts [22].

A complementary and underused source of early cognitive signals is unstructured clinical documentation, including HHC nursing notes [10,23]. Recent reviews show that natural language processing (NLP) approaches applied to electronic health record (EHR) notes can identify cognitive impairment with strong median performance across studies, but variability in diagnostic criteria, data sources, and external validation remains a key barrier to translation [24]. In parallel, large language model (LLM) methods are increasingly being evaluated for detecting cognitive decline from clinical notes, including large clinical language model approaches (eg, CD-Tron) and comparative studies of LLMs in real-world clinical text [25-29]. These developments suggest that LLM-enabled extraction can help capture both explicit and implicit mentions of symptoms, risk factors, and functional concerns that are inconsistently represented in structured fields, an especially relevant issue in home-based care workflows.

HHC is therefore a compelling setting for scalable, equity-oriented screening because it provides repeated encounters and routinely generates multiple complementary data streams, including standardized assessments (eg, Outcome and Assessment Information Set [OASIS] in US Medicare–certified home health agencies), narrative nursing notes, and patient-nurse verbal communication. Our prior work in HHC has shown that combining structured assessment data with information extracted from clinical notes can improve risk identification (HomeADScreen [12]). More recently, we demonstrated the potential value of leveraging audio-recorded patient-nurse verbal communication as an additional signal beyond EHR data for early cognitive screening in HHC [16]. However, few studies have jointly leveraged (1) standardized home-care assessments, (2) unstructured home-care clinical notes, and (3) routine patient-clinician conversations within a single integrated screening framework for mild cognitive impairment and early dementia (MCI-ED) in home-based care.

Accordingly, this study describes an ongoing protocol to develop and evaluate a multimodal screening approach for identifying MCI and early dementia in HHC using routinely generated data streams: standardized assessment data (OASIS), HHC nursing notes, and audio-recorded patient-nurse verbal communication. We aim to (1) model speech, language, emotion, and interaction

patterns from patient-nurse conversations using automated speech analysis, (2) apply NLP/LLM methods to identify MCI/early dementia–related symptoms, lifestyle risk factors, and communication deficits from both clinical notes and verbal communication, and (3) integrate these signals with standardized assessment variables to improve screening performance compared with models based on any single data stream.

Methods

Study Setting, Design, and Status

This protocol describes an ongoing cross-sectional case-control protocol in collaboration with VNS Health (formerly Visiting Nurse Service of New York), one of the largest HHC systems in the United States. The study population includes adults aged 60 years and older who receive HHC services from VNS Health. The protocol is designed to develop and evaluate a multimodal screening algorithm for identifying MCI-ED in HHC.

Participant Recruitment, Eligibility, and Group Allocation

Recruitment Focus and Rationale

Aim 1 focuses on modeling speech, language, emotional expression, and interaction patterns during routine patient-nurse encounters (verbal communication) as markers of early cognitive decline in HHC. The primary analytic cohort includes non-Hispanic Black and non-Hispanic White patients receiving HHC services from VNS Health. These groups were selected because they are highly represented in the study setting, enable adequately powered comparisons within a single HHC system, and facilitate evaluation of model performance across racial groups—particularly important given well-documented disparities in dementia diagnosis and care for Black patients.

Recruitment Strategy

Potential participants are identified through EHR-based screening and clinician referral workflows within VNS Health. Prespecified EHR indicators are used to identify likely cases (eg, documented symptoms of cognitive decline) and likely controls (no evidence of impairment). Eligible patients are approached during an active episode of HHC. Recruitment is monitored to achieve representation of both racial groups and, when feasible, balance across key characteristics, age, sex as a biological variable, and education.

Eligibility, Screening Workflow, and Group Allocation

Eligible participants are aged ≥60 years, plan to receive VNS Health services during the study period, have sufficient English proficiency to communicate independently with HHC nurses, have adequate vision/hearing to complete cognitive testing, and can provide written informed consent. Patients are excluded if they (1) are unable to communicate independently with the HHC nurse in English and (2) have speech or language disorders due to neurological conditions other than MCI-ED (eg, Parkinson disease or seizure disorders). Full eligibility criteria are provided in Multimedia Appendix 1.

We use a 2-stage approach to identify patients for case and control groups. First, we identify potential cases using available

ICD-10 (*International Statistical Classification of Diseases, Tenth Revision*) diagnoses in the EHR (*ICD-10* G31.84 for MCI) and identify potential controls as patients without documented cognitive impairment. Second, all consented participants without an existing MCI diagnosis complete cognitive assessments—the Montreal Cognitive Assessment (MoCA) [30,31] and Clinical Dementia Rating (CDR) [31]—in their homes, administered by a trained research assistant who audio-records responses to support final group assignment.

A study clinician with expertise in cognitive impairment detection reviews the recorded cognitive assessments together with relevant clinical context (medical history and nurse assessment information from OASIS) to confirm group assignment. Based on prespecified criteria, participants are classified as MCI-ED cases when findings are consistent with early cognitive impairment (anticipated CDR 0.5-1 and MoCA ~16-25, with consideration of EHR evidence when available) and as cognitively normal controls when findings are within normal limits (CDR 0 and MoCA ≥26), and there is no EHR evidence of cognitive impairment. Participants meeting criteria for moderate to severe impairment (eg, CDR 2-3 or MoCA <16) are excluded because the protocol focuses on MCI-ED.

After group allocation, patients in both the case and control groups are invited to provide additional consent for the next phase of Aim 1, which includes audio-recording routine patient-nurse encounters.

Ethical Considerations

This study was reviewed and approved as human participant research by the Columbia University Irving Medical Center Institutional Review Board (Protocol AAAU3168). The study is conducted in collaboration with VNS Health and complies with all applicable institutional, federal, and regulatory requirements for research involving human participants.

Written informed consent is obtained from all participating patients prior to enrollment. Consent includes permission for the administration of cognitive assessments, including the MoCA and the CDR, audio-recording of patient-nurse encounters, and linkage of audio recordings and assessment data with EHR information. HHC nurses also provide informed consent for participation, including consent for audio-recording of patient-nurse encounters. Participants are informed of the study purpose, procedures, potential risks, and their right to withdraw at any time without affecting their care or employment.

All study data are handled in accordance with Health Insurance Portability and Accountability Act and institutional data protection policies. Audio recordings, transcripts, cognitive assessment data, and clinical text are deidentified prior to analysis, with direct identifiers removed. Data are stored on secure, access-controlled servers at Columbia University and VNS Health with role-based permissions and audit logging. Access to identifiable data is restricted to authorized study personnel only. Deidentified datasets are used for analysis, and results are reported in aggregate to minimize the risk of participant reidentification.

Each participating patient receives a US $50 incentive for completion of cognitive assessments (MoCA and CDR) and an additional US $50 incentive for participation in audio-recording of patient-nurse encounters. HHC nurses also receive a US $50 incentive for participation in audio-recording of patient-nurse encounters. Incentives are provided in accordance with institutional review board–approved procedures and are not contingent on study outcomes.

Data Collection Overview

Number of Audio-Recorded Encounters (Patient-Nurse Conversation)

For each enrolled participant, 3 routine patient-nurse encounters are audio-recorded during the HHC episode of care. Recording multiple encounters provides repeated observations to capture within-person variability in speech, language, and interactional patterns across visits, while minimizing participant and clinician burden. When both patient and nurse provide consent, a trained research assistant attends the visit and operates a Saramonic Blink audio-recording device [32], minimizing burden on clinical staff. This portable device, with dual wireless microphones that attach to clothing, provides clear speech transmission to devices like an iPod and offers dual-channel storage.

Linking Audio-Recorded Encounters to EHR Data and Clinical Notes

Audio-recorded encounters are linked to EHR data extracted from the VNS Health system, including the OASIS [33,34]—a federally mandated HHC assessment capturing patient health status, functional status, and living arrangements—as well as supplemental structured data (eg, medications) and free-text clinical notes. Free-text notes include visit notes, documenting each nurse encounter, and care coordination notes, capturing communications with other clinicians, physicians, and family members.
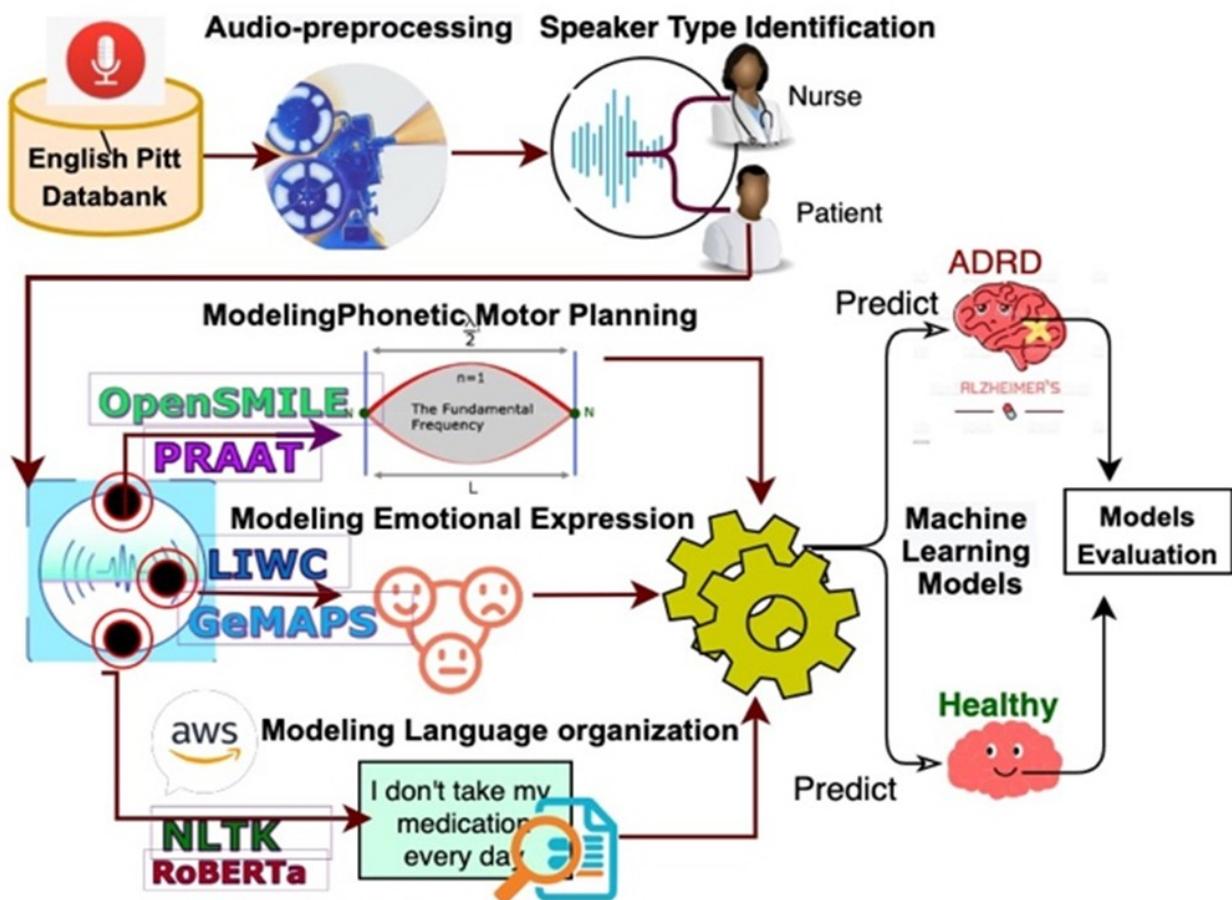
Preliminary Feasibility and Pilot Studies

Prior to this protocol, we conducted a series of pilot studies to establish the feasibility of audio-recording patient-nurse verbal communication and applying automated speech and machine learning methods in the HHC setting [22]. First, we evaluated several commercially available audio-recording devices in laboratory and real-world HHC settings, assessing usability, transcription quality, and acceptability among HHC nurses and patients. Based on System Usability Scale scores and transcription accuracy measured by word error rate, the Saramonic Blink device [35] demonstrated the best overall performance and was selected for use in the current study. Semistructured interviews with HHC nurses and patients further indicated that audio-recording was acceptable and had minimal perceived impact on routine care delivery.

In a second pilot study, we demonstrated the feasibility of automated speaker type identification in recorded HHC encounters using machine learning models trained on acoustic and lexical features, achieving satisfactory classification performance [36]. In a third pilot study, we developed and validated an end-to-end analytic pipeline for modeling spoken language in cognitive impairment [11] (Figure 1). The pipeline included (1) audio preprocessing for noise reduction; (2) automated speaker type identification to separate patient and clinician speech; (3) extraction of acoustic features capturing phonetic motor planning and voice characteristics (eg, fluency, frequency/spectral measures, intensity, and instability) using OpenSMILE [37] and PRAAT [38]; (4) modeling of emotional expression using the Geneva Minimalistic Acoustic Parameter Set [39] (GeMAPS) complemented by lexicon-based psycholinguistic markers [40] (using Linguistic Inquiry and Word Count [LIWC]); (5) modeling of language organization using transcript-derived lexical and syntactic measures—using Natural Language Toolkit (NLTK) [41]—and contextual language representations (using distilled RoBERTa [42]); and (6) machine learning–based classification with internal validation. We evaluated this pipeline on a benchmark dataset (DementiaBank [43] "Cookie Theft" picture descriptions) and observed strong discrimination between cognitively impaired and cognitively unimpaired participants, supporting the feasibility of extracting informative speech-derived markers and training predictive models. Collectively, these pilot studies informed the design decisions, data collection procedures, and analytic pipelines in this protocol.

Figure 1. Analytic pipeline for modeling spoken language. ADRD: Alzheimer disease and related dementias; LIWC: Linguistic Inquiry and Word Count; NLTK: Natural Language Toolkit.

Analytic Method for Aim 1: Model MCI-ED Patients' Verbal Communications With Clinicians Using an Automated Speech Analysis System

Rationale and Overview

Early cognitive decline affects multiple aspects of spoken communication, including speech motor control, language organization, emotional expression, and social interaction. In HHC settings, these changes are expressed during spontaneous patient-nurse verbal communications rather than structured speech production tasks (eg, reading task). Building on our preliminary feasibility and pilot work, Aim 1 focuses on systematically modeling these communication patterns using an automated speech analysis system. The objective is to extract complementary acoustic, linguistic, emotional, and interactional features from naturally occurring patient-nurse communications that may signal MCI-ED. The analytic framework for Aim 1 consists of 5 components, summarized in Table 1, which together capture core dimensions of speech production and interaction relevant to cognitive decline.

Table 1. Modeling mild cognitive impairment and early dementia (MCI-ED) patient-nurse verbal communication in the home health care setting (components 1-4).

Component and domain	Measures
Component 1: modeling phonetic motor planning	
Speech (vocal) fluency	• Articulation: number of phonemes per second without hesitation [44].
	• Speech rate: number of phonemes per second with hesitation [44].
	• Silent pauses: number of speechless intervals at the beginning of and between words [45].
	• Within-word disfluency: within-word silent pauses and sound prolongations [46].
Rhythmic structure of speech	• Syllabic intervals: temporal variability in speech [47].
	• Pairwise variability index: durational variability in successive acoustic-phonetic intervals [48].
	• Vowel duration: proportion of time of vocalic intervals in a sentence and the standard deviation of inter-vowel intervals [49].
Frequency and spectral domain	• Fundamental frequency: average number of oscillations originating from the vocal folds per second [50].
	• Formant frequencies (F1-F4): acoustic resonances of the vocal tract due to changes in the positions of vocal organs [51].
	• Spectral center of gravity: amplitude-weighted mean of harmonic peaks averaged over sound duration [52].
	• Long-term average spectrum: composite signal representing the spectrum of the glottal source and resonant characteristics of the vocal tract [53].
	• Mel-frequency cepstral coefficients: energy variations between frequency bands of a speech signal [54].
Voice instability	• Jitter: cycle-to-cycle period variation of successive glottal cycles [55].
	• Shimmer: cycle-to-cycle amplitude variation of successive glottal cycles [55].
	• Cepstral peak prominence: measure of periodicity in the speech signal [56].
Voice quality	• Harmonics-to-noise ratio: relative amount of additive noise in the voice signal [50].
	• Voice breaks: reduced ability in vocal cord execution resulting in voice breaks [57].
	• Acoustic voice quality index: weighted combination of time-frequency and quefrency-domain metrics developed to measure the severity of dysphonia [58].
Voice intensity	• Hammarberg index: articulatory effort computed as the difference between maximum energy in the 0-2 kHz band and the energy in the 2-5 kHz band [59].
	• Energy concentration: average spectral frequency [45].
Component 2: modeling the patient's emotional expression	
Frequency parameters	• Pitch: number of vibrations per second produced by the vocal cords [60].
	• Jitter [55].
	• Center frequency of formants 1-3.
	• Bandwidth of formants 1-3. Formant frequencies are acoustic resonances of the vocal tract caused by changes in vocal organ positions [51].
Energy/amplitude	• Shimmer.
	• Loudness: estimate of perceived signal intensity from an auditory spectrum [61].
	• Harmonics-to-noise ratio: relative amount of additive noise in the voice signal [50].
Spectral parameters	• Alpha ratio: ratio of summed energy from 50-1000 Hz and 1-5 kHz.
	• Hammarberg index: measure of articulatory effort [59].
	• Spectral slope (0-500 Hz and 500-1500 Hz).
	• Formant 1-3 relative energy.
	• Harmonic difference H1-H2: difference between first and second harmonic amplitudes [62].
	• Harmonic difference H1-A3: difference between H1 and A3 (energy of the highest harmonic in the third formant range) [62].
	• Spectral flux: difference between the spectra of 2 consecutive frames.
	• 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	• Moving average type-token ratio: total number of unique words divided by the total number of words for each successive fixed-length window [63].
	• Brunet index: variation in word types marked by part-of-speech tagging relative to the total number of words in a sentence [64].
	• Honore index: proportion of words used only once relative to the total number of words [65].

Component and domain	Measures
Syntactic level of language organization	• Sentence complexity: score computed using a syntactic parse tree [66]. • Grammatical errors: identified using a parse tree analyzer [67]. • Incomplete (fragment) sentences: identified using an automatic detection algorithm based on syntactic parse trees and part-of-speech tagging [68].
Semantic fluency	• Identification of filled pauses (eg, "um") in the patient's spoken language [69-72].
Patient recall ability	• Uncertainty in patient language: computed using the linguistic approximator introduced by Ferson et al [73]. • Memory-related terms: proportion of sentences containing memory-related terms relative to the total number of sentences, computed using the NimbleMiner toolkit [74,75]. • Question ratio: proportion of interrogative sentences relative to the total number of sentences, identified using the NLTK[a] Python package [74].
Component 4: modeling patient-nurse interaction	
Patient turns	• Continuous block of uninterrupted speech by a single patient. • Total number of patient turns indicates frequency of information exchange [76].
Interactivity	• Dialog interactivity: defined as the total number of patient turns divided by the total length of the encounter [76].
Turn density	• Computed using the same parameters specified for lexical richness (component 3).
Turn duration	• Length of time of the patient's turn; longer durations have been associated with difficulty in turn monitoring in MCI-ED [76].
Relative timing of turns	• Discernible pause rate: proportion of discernible speechless intervals at the start of patient turns relative to total utterances [77]. • Cross-over speaking rate: proportion of patient-nurse utterances with cross-over speaking relative to total utterances during the interaction [77].

[a]NLTK: Natural Language Toolkit.

Feature Specification and Reproducibility

For reproducibility, all speech- and interaction-based parameters extracted in Aim 1 are explicitly specified in Table 1, organized by analytic component (phonetic motor planning, emotional expression, syntactic and semantic language organization, and patient-nurse interaction). Table 1 provides the operational definition and measurement domain for each parameter. Acoustic features are computed using established toolkits (OpenSMILE [78] and PRAAT [38]); linguistic features are derived from automatically transcribed speech using NLTK, LIWC [79], and distilled RoBERTa; and interactional features are computed from speaker-labeled timestamps generated by Amazon Web Services (AWS) Transcribe.

Component 1: Modeling Phonetic Motor Planning

Impairment in phonetic motor planning is a well-documented consequence of neurodegenerative disorders, including MCI-ED, and manifests as reduced articulation precision, altered speech rhythm [44,80,81], and increased disfluency. To characterize these changes, we analyze acoustic parameters across six domains (Table 1, component 1): (1) speech fluency [45,46,82], (2) rhythmic structure [48,49,83], (3) frequency and spectral characteristics [45,84,85], (4) voice instability [45,69,86], (5) voice quality [87-89], and (6) voice intensity [45,90]. These measures quantify temporal and spectral aspects of speech that reflect the patient's ability to plan and execute vocal motor actions.

Component 2: Modeling Emotional Expression

Alterations in emotional expression often develop alongside cognitive decline and can negatively affect communication quality and interpersonal interaction [91,92]. Emotion is conveyed both through nonverbal vocalization and semantic content [93-96]. To model vocal expression of emotion, we use the GeMAPS [39], which captures affect-related changes in autonomic arousal and vocal musculature via frequency-, energy-, and spectral-domain parameters (Table 1, component 2). To capture the semantic expression of emotion, we extract linguistically encoded emotional indicators using the LIWC dictionary. Emotion-related linguistic markers (eg, sadness and anxiety) have been associated with cognitive dysfunction [97] and adverse health outcomes [98].

Component 3: Modeling Syntactic, Semantic, and Pragmatic Language Organization

Language impairment in MCI-ED is characterized by reduced lexical diversity, simplified syntax [44,97,99], word-finding difficulties [44,100], and impaired memory-related discourse [44,101], which together contribute to reduced coherence. We model language organization using features in four domains (Table 1, component 3): (1) lexical richness [64,65,102], (2) syntactic complexity and grammaticality [66,103], (3) semantic fluency [104], and (4) patient recall ability [73,74]. Linguistic features are derived from automatically transcribed speech and processed using standard NLP tools. In addition, we will use

distilled RoBERTa [42] to generate contextual language representations that capture semantic relationships beyond surface-level lexical features.

Component 4: Modeling Patient-Nurse Interaction

Cognitive impairment also affects social communication, including turn-taking, timing, and responsiveness [105,106]. Patients with MCI-ED may show recurrent interactional patterns, such as longer turns, delayed responses, or reduced interactivity [107]. To capture these phenomena, we will model patient-nurse interaction using easily measurable dialogue features [76,77,108] (Table 1, component 4), including patient turn counts, dialog interactivity, turn density, turn duration, and relative timing of turns. These interactional measures reflect how patients engage with clinicians in real-world care encounters and provide information beyond speech content alone.

Component 5: System Implementation and Feature Extraction

Acoustic parameters for components 1 and 2 will be computed at the utterance level (continuous blocks of uninterrupted patient speech) using OpenSMILE [37] and PRAAT toolkits [38]. Verbal communications are automatically transcribed using AWS Transcribe, after which linguistic features for component 3 are computed at the encounter level using the NLTK toolkit and distilled RoBERTa. Interactional features for component 4 are derived from AWS Transcribe metadata, including speaker labels and time stamps. All extracted features will be aggregated at the patient level for downstream integration with clinical data in Aim 3.

Analytic Method for Aim 2: Extraction of MCI-ED–Related Information From Clinical Text Using LLMs

Rationale and Overview

Many clinical indicators of MCI-ED—including symptoms, lifestyle risk factors, and communication difficulties—are documented in free-text clinical notes or expressed during patient-nurse conversations [109], but are not captured in structured EHR fields [6,110]. Aim 2 uses LLMs to systematically extract this information from HHC clinical notes and transcripts of patient-nurse verbal communication. The goal is to convert unstructured text into standardized, patient-level variables that can be integrated with speech features (Aim 1) and structured assessment data (OASIS) for multimodal screening (Aim 3).

Information Specification and Reproducibility

For reproducibility, all MCI-ED–related information identified in Aim 2 is defined using an information schema summarized in component 1 (Information targets and schema). The schema specifies the target information families (clinical symptoms, lifestyle risk factors, and communication deficits) and associated attributes, and is applied consistently across human annotation and LLM-based identification. MCI-ED–related information is normalized to standard clinical terminologies—Unified Medical Language System (UMLS) [111] concepts, when available—and represented in a structured patient-level format, supporting reproducible integration with Aim 1 features and OASIS variables.

Component 1: Information Targets and Schema

We define a schema for the 3 MCI-ED–related risk factor categories, including clinical symptoms, lifestyle risk factors, and communication deficits. For each identified item, the system records (1) the related terms; (2) a normalized clinical concept identifier when available (UMLS [111] concepts); (3) clinically relevant attributes: assertion (present/absent/possible), temporality (current/historical), and experiencer (patient/caregiver); (4) severity and frequency; and (5) duration. The schema is used consistently across clinical notes and transcripts of patient-nurse communication.

Component 2: Human-Annotated Gold-Standard Dataset

Using the information schema defined in component 1, we create a human-annotated gold-standard dataset to support LLM adaptation and evaluation. The schema specifies the target information categories (clinical symptoms, lifestyle risk factors, and communication deficits) and associated attributes, including assertion, temporality, and experiencer.

Two trained nurse annotators independently annotate a stratified sample of HHC clinical notes and encounter transcripts according to this predefined schema. The annotation sample is stratified by race, sex, and visit type to ensure representation of diverse documentation patterns. Interannotator agreement is assessed using Cohen κ [112] for each information category and attribute, calculated on double-annotated samples prior to adjudication. Discrepancies are resolved through adjudication meetings to produce a finalized gold-standard dataset. The annotated corpus is subsequently partitioned into training, development, and test sets to enable LLM fine-tuning and unbiased performance evaluation.

Component 3: LLM-Based Identification Strategy

We use a hybrid LLM-based strategy that combines prompted extraction and instruction tuning [113], both aligned with the information schema (component 1) and supervised by the human-annotated gold-standard dataset (component 2). First, prompted extraction (baseline): as a baseline approach, we apply structured prompts that explicitly define the 3 information families (clinical symptoms, lifestyle risk factors, and communication deficits) and required attributes (assertion, temporality, and experiencer). Prompts instruct the model to produce schema-compliant JSON outputs and to provide a supporting text span for each identified item to ensure evidence-grounded extraction. This prompted approach provides an interpretable, rapidly adjustable method for early experiments and error analysis. Second, instruction tuning (schema-guided supervised adaptation): to improve reliability on HHC-specific language and documentation patterns, we perform instruction tuning using the training split of the human-annotated gold-standard dataset. Training examples pair the input text (note or transcript segment) with the target output formatted as schema-compliant JSON, including the identified item type, attributes (assertion, temporality, and experiencer), and supporting span. This teaches the model to follow the extraction instructions consistently and to produce outputs that match the

schema across diverse note styles and conversational phrasing. Instruction tuning is implemented using parameter-efficient methods (eg, Low-Rank Adaptation [114]/ Quantized Low-Rank Adaptation [114,115]) to reduce computational burden in the secure environment.

Component 4: Normalization and Patient-Level Aggregation

For each identified item in the text, we normalize the item (mention) to a standardized clinical concept identifier (UMLS, when available) using a controlled vocabulary lookup supplemented by string similarity matching for common variants. When multiple items map to the same concept within a document, we merge them into a single record while preserving the schema attributes (assertion, temporality, experiencer, and—when present—severity and frequency/duration) and retaining the supporting text spans for traceability. We then aggregate document-level outputs to the patient level to produce predictors for Aim 3. Patient-level variables summarize the presence of each normalized concept and its attributes across the patient's available HHC notes and encounter transcripts. The final deliverable is a structured patient-level table of normalized MCI-ED–related symptoms, lifestyle risk factors, and communication deficits for integration with OASIS data and Aim 1 speech and interaction features.

Component 5: Evaluation, Subgroup Checks, and Quality Control

We evaluate performance against the human-annotated gold-standard dataset using (1) span-level precision/recall/F_1 under exact and overlap matching, (2) attribute performance (assertion, temporality, experiencer, and severity/frequency when applicable), and (3) concept normalization accuracy. We report results overall and stratified by race, sex, and age group to monitor for systematic performance differences. We conduct routine error analysis (eg, common false positives from templated note language, negation errors, or transcript artifacts) and use findings to refine prompts, update normalization resources, and adjust fine-tuning settings. Low-confidence outputs are flagged for targeted review during development to guide iteration and reduce systematic errors.

Component 6: Aim 2 Outputs Used in Aim 3

The final Aim 2 deliverable is a structured, patient-level dataset summarizing clinical symptoms, lifestyle risk factors, and communication deficits identified from HHC notes and transcripts, including normalized concept identifiers and clinically meaningful attributes. These variables are used as candidate predictors and complementary signals in the multimodal screening algorithm developed in Aim 3.

Analytic Method for Aim 3: Development of a Multimodal Screening Algorithm for Identifying MCI-ED in HHC

Objective and Overview

The objective of Aim 3 is to develop and evaluate a multimodal screening algorithm for identifying HHC patients with MCI-ED. The algorithm integrates complementary information from three routinely generated data sources: (1) speech and interaction features extracted from patient-nurse communication (Aim 1), (2) MCI-ED–related information identified from clinical notes and transcripts (Aim 2), and (3) structured assessment data from the OASIS.

Data Sources and Candidate Predictors

Input variables include (1) acoustic, linguistic, emotional, and interactional features derived from patient-nurse verbal communication (Aim 1); (2) normalized clinical symptoms, lifestyle risk factors, and communication deficits identified from clinical notes and encounter transcripts (Aim 2); and (3) structured OASIS variables capturing sociodemographic characteristics, diagnoses, medications, functional status, and related clinical information.

Component 1: Data Preprocessing and Feature Preparation

Prior to model development, we assess data quality and address missingness, inconsistency, and integrity issues using a predefined data quality framework [116]. Continuous variables are transformed and scaled as appropriate to ensure comparability across modalities. To reduce dimensionality and mitigate overfitting in the presence of a large number of candidate predictors, we apply Joint Mutual Information Maximization [117] as a feature selection method. Joint Mutual Information Maximization is selected for its suitability in small to moderate sample settings with high-dimensional data, where it balances relevance to the outcome with redundancy among features.

Component 2: Model Development and Multimodal Integration

We develop screening models using supervised discriminative machine learning algorithms that are appropriate for tabular and multimodal clinical data, including logistic regression, support vector machines (SVMs) [118], and ensemble tree-based methods [119-122]. These models are chosen for their interpretability, robustness, and reduced risk of overfitting in clinical datasets. Multimodal integration is performed by combining features from speech, clinical text, and OASIS data within a unified modeling framework. Models are trained to estimate the probability of MCI-ED at the patient level. Temporal aspects of speech-derived features are summarized at the patient level prior to modeling, rather than modeled using complex sequence architectures, to maintain feasibility and stability given sample size considerations.

Component 3: Model Training, Validation, and Fairness Assessment

Model training and hyperparameter tuning are conducted using stratified nested cross-validation, with inner loops for parameter selection and outer loops for performance estimation. Model performance is evaluated using the area under the receiver operating characteristic curve (AUC-ROC) and area under the precision-recall curve, along with calibration measures to assess agreement between predicted risk and observed outcomes. To evaluate equitable performance, we assess model metrics stratified by race, sex, and age group. Fairness-related measures,

including group-wise differences in sensitivity and specificity and calibration across subgroups, are examined. When systematic performance differences are observed, we explore mitigation strategies such as reweighting or threshold adjustment and reassess model performance.

Component 4: Final Model Evaluation and Output

In the final step, the selected model is evaluated on an independent validation dataset to provide an unbiased estimate of performance. The screening algorithm produces a patient-level risk score indicating the likelihood of MCI-ED, which is intended to support clinical awareness and referral for further cognitive evaluation rather than serve as a diagnostic tool. The resulting algorithm and associated feature sets are prepared for downstream evaluation of clinical use and integration into HHC workflows.

Results

Initial Study Population and Clinical Characteristics

Between February 2024 and July 2025, we enrolled 114 HHC patients who met eligibility criteria and completed study-administered cognitive assessments. Following standardized review of the cognitive assessments and prespecified group-allocation procedures, 55 participants were classified as MCI-ED cases and 59 as cognitively normal controls. The cohort had a balanced sex distribution (n=58, 51% female) and was racially and ethnically diverse (n=63, 55.3% Black). Most participants were insured by Medicare (n=75, 66%), and 44% (n=50) lived alone.

In descriptive comparisons, participants classified as cognitively impaired had a higher prevalence of urinary incontinence (21/55, 36.8% vs 13/59, 21.4%), anxiety (32/55, 57.9% vs 23/59, 39.3%), and impaired vision (14/55, 26.3% vs 4/59, 7.1%), as well as greater dependence in activities of daily living (41/55, 73.7% completely dependent vs 38/59, 64.3%). These results are reported to characterize the cohort and should be interpreted descriptively rather than as definitive group differences.

Audio-recorded patient-nurse encounters had a median duration of 19 (IQR 12-23) minutes and a median of 56 (IQR 31-80) utterances per encounter. Across encounters, nurses contributed more words than patients (median 842, IQR 461-1218 vs 589, IQR 303-960), consistent with the structure of routine HHC visits and motivating inclusion of interactional features.

Preliminary Modeling Results

We conducted exploratory modeling analyses to evaluate the feasibility of distinguishing MCI-ED cases from cognitively normal controls using (1) speech-derived measures, (2) clinical text, and (3) structured EHR/OASIS variables, as well as multimodal combinations of these data sources. These analyses are intended to assess feasibility and inform subsequent model refinement and validation, rather than to provide definitive estimates of performance.

Speech-Derived Representations

Acoustic and temporal speech features were encoded using SpeechDETECT, including parameters related to phonetic motor planning (Table 1, component 1). Vocal emotion-related cues were encoded using GeMAPS (Table 1, component 2). Linguistic features included handcrafted measures capturing lexical richness, syntactic complexity, and semantic/fluency markers (eg, repetition and filler words; Table 1, component 3), and psycholinguistic indicators were extracted using LIWC 2015. In addition, we evaluated pretrained transformer language models for transcript-based representations.

Unimodal Performance

When modeling patient speech alone, DistilBERT achieved the strongest performance among evaluated BERT-based models (F_1=69.39; AUC-ROC=69.36). For clinical notes, BioClinicalBERT yielded the best performance among evaluated language models (F_1=64.29; AUC-ROC=69.17). Among traditional classifiers, a linear SVM performed well using patient speech features (F_1=75.0; AUC-ROC=75.94). Models using structured EHR/OASIS variables achieved their best performance with logistic regression (F_1=75.56; AUC-ROC=79.70).

Nurse Speech and Interactional Features

Incorporating nurse speech and interactional measures (Table 1, component 4) resulted in improved discrimination (SVM F_1=85.0; AUC-ROC=86.47), suggesting that patient-nurse interaction captures complementary information beyond patient speech alone.

Multimodal Integration

In multimodal analyses integrating speech features, interactional measures, and structured EHR/OASIS variables, the SVM achieved the highest overall performance (F_1=88.89; AUC-ROC=90.23). Examination of model contributions suggested that reduced lexical diversity, longer patient pauses, increased nurse dominance in conversation, selected psycholinguistic markers, and specific EHR variables (eg, non–insulin-dependent diabetes, pressure ulcers, and living alone) contributed to discrimination.

Overall, these results support the feasibility of extracting and integrating multimodal signals for MCI-ED screening in HHC. Final model evaluation, subgroup performance assessment, and fairness analyses will be conducted using the prespecified validation procedures after completion of recruitment and the finalized analytic dataset.

Discussion

Overview

This study protocol describes a multimodal screening approach for identifying MCI-ED in HHC using routinely generated data streams. The central hypothesis is that spontaneous patient-nurse conversations, combined with structured HHC assessment data (the federally mandated OASIS instrument) and information extracted from free-text clinical documentation, can provide complementary signals for earlier identification of cognitive impairment than any single data stream alone.

The implementation results reported in this manuscript demonstrate the feasibility of an end-to-end workflow in HHC,

XSL•FO

RenderX

including audio capture during routine visits, automated transcription and speaker labeling, extraction of acoustic, linguistic, emotional, and interactional features, and linkage to clinical notes and OASIS variables. Exploratory analyses in the analyzed cohort suggest that incorporating nurse speech and interactional features can improve discrimination beyond patient speech alone, consistent with the premise that conversation structure (eg, timing, pauses, turn-taking balance, and interactivity) contains clinically relevant information. These findings should be interpreted as feasibility and proof-of-concept evidence, rather than definitive estimates of model performance.

A substantial body of prior work has demonstrated that speech markers can differentiate individuals with Alzheimer disease from cognitively unimpaired controls, often using structured or semistructured tasks (eg, picture description, verbal fluency, and reading) collected in controlled environments [13]. While these approaches have been valuable for benchmarking and understanding underlying patterns, they may be less sensitive to the subtle and heterogeneous manifestations of MCI-ED and may not reflect communication behaviors during real-world clinical encounters.

This study extends this literature in 3 important ways. First, it shifts the speech signal from standardized tasks to naturally occurring clinical interactions in HHC, where pragmatic, temporal, and turn-taking patterns can be observed at scale. Second, it models not only patient speech characteristics but also interactional dynamics (including nurse speech), which may reflect clinician adaptation to support patients and/or patient difficulty maintaining conversational flow. Third, it advances a multimodal framework by integrating conversational speech features with (1) structured assessment variables from OASIS and (2) MCI-ED–related information embedded in free-text documentation or spoken conversation but not consistently represented in structured EHR fields. Together, these extensions aim to improve the practical relevance of screening in HHC settings where comprehensive cognitive evaluations may be limited.

Recent advances in minimally invasive approaches to Alzheimer disease detection, including blood-based biomarker testing in symptomatic individuals, reflect increasing clinical emphasis on earlier identification. However, biomarker confirmation alone does not characterize how cognitive decline affects communication during routine care. Changes in speech and language, such as reduced fluency, disrupted discourse organization, and altered vocal control, often emerge early and reflect functional impairment that is not captured by biological measures. The screening approach described here targets this complementary dimension by modeling communication behaviors observed in routine patient-nurse encounters, providing ecologically valid indicators of cognitive change. Integrating speech-based indicators with other clinical information, including biomarker evidence when available, may support a more comprehensive assessment of cognitive decline and its impact on real-world functioning.

Limitations

Several limitations should be considered when interpreting these results. First, findings are based on data from a single HHC organization and a modest sample, which may limit generalizability and yield performance estimates that are sensitive to sampling variability. Second, audio quality, background noise, and automated transcription/speaker-labeling errors can affect the accuracy of extracted acoustic, linguistic, and interactional features. Third, interactional measures may reflect both patient cognitive-linguistic status and clinician communication style or workflow constraints, which can introduce confounding if not explicitly modeled. Fourth, interactional measures may reflect both patient cognitive-linguistic status and clinician communication style or workflow constraints, which can introduce confounding if not explicitly modeled. Fourth, the protocol focuses on English-speaking participants with sufficient hearing/vision to complete cognitive testing; results may not generalize to other language groups or to patients with sensory limitations, who are common in HHC. Finally, cross-sectional classification does not establish whether speech and interaction markers predict future cognitive trajectories, underscoring the need for longitudinal evaluation.

Future Directions

Several avenues for future research could strengthen both the scientific rigor and clinical use of this approach. First, prospective longitudinal studies are needed to move beyond cross-sectional classification and evaluate whether speech and interactional markers can predict cognitive decline trajectories or functional deterioration over time. Such studies would clarify whether these features capture progressive change in addition to baseline differences. Second, external validation across diverse HHC agencies, geographic regions, and care delivery models will be essential to assess model transportability and identify when recalibration is necessary. Third, more granular analysis of conversational dynamics could distinguish clinically meaningful interaction patterns—such as repair sequences, prompting behaviors, and topic maintenance difficulties—from structural features that primarily reflect workflow or documentation practices. Fourth, incorporating clinician feedback through human-in-the-loop development cycles can help identify model failure modes, enhance interpretability of predictions, and establish safe deployment thresholds informed by real-world use cases. Finally, pragmatic clinical use trials are needed to determine whether integrating speech-based screening into HHC workflows improves downstream outcomes, including timeliness of formal cognitive evaluation, care plan modifications, and patient safety. Collectively, these efforts would bridge the gap between technical performance and meaningful improvements in care delivery for older adults at risk of cognitive decline.

Dissemination Plan

Findings will be disseminated through peer-reviewed publications and presentations to clinical and informatics audiences. To support reproducibility while protecting privacy, the study team plans to share (1) detailed feature definitions and extraction procedures, (2) deidentified analytic code and configuration files where permissible, and (3) the annotation schema and evaluation framework for text extraction.

Acknowledgments

Funding

Conflicts of Interest

None declared.

Multimedia Appendix 1

Eligibility criteria.
[DOCX File , 17 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Peer review report from AGCD-1 - Career Development Facilitating The Transition to Independence Study Section, National Institute on Aging (National Institutes of Health, USA).
[PDF File (Adobe PDF File), 127 KB-Multimedia Appendix 2]

References

1. Dementia. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/dementia?utm_source=chatgpt.com [accessed 2025-12-16]
2. World Alzheimer Report 2025. Alzheimer's Disease International. URL: https://www.alzint.org/resource/world-alzheimer-report-2025/ [accessed 2025-12-17]
3. 2025 Alzheimer's disease facts and figures. The Journal of the Alzheimer's Association. 2025;21(4):e70235. [doi: 10.1002/alz.70235]
4. Livingston G, Huntley J, Liu KY, Costafreda SG, Selbæk G, Alladi S, et al. Dementia prevention, intervention, and care: 2024 report of the Lancet Standing Commission. Lancet. 2024;404(10452):572-628. [doi: 10.1016/S0140-6736(24)01296-0] [Medline: 39096926]
5. Zolnour A, Azadmaleki H, Haghbin Y, Taherinezhad F, Nezhad MJM, Rashidi S, et al. LLMCARE: early detection of cognitive impairment via transformer models enhanced by LLM-generated synthetic data. Front Artif Intell. 2025;8:1669896. [FREE Full text] [doi: 10.3389/frai.2025.1669896] [Medline: 41280882]
6. Barrón Y, Ryvicker M, Song J, Zolnoori M, Topaz M. Identifying new ADRD diagnoses in home health care patients using natural language processing of nurses' notes. Innov Aging. 2023. 2023;7(Suppl 1):1060. [doi: 10.1093/geroni/igad104.3406]
7. Song J, Zolnoori M, McDonald MV, Barrón Y, Cato K, Sockolow P, et al. Factors associated with timing of the start-of-care nursing visits in home health care. J Am Med Dir Assoc. 2021;22(11):2358-2365. [FREE Full text] [doi: 10.1016/j.jamda.2021.03.005] [Medline: 33844990]
8. 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. [FREE Full text] [doi: 10.1177/2333721420959861] [Medline: 33029550]
9. Ryvicker M, Barrón Y, Shah S, Moore SM, Noble JM, Bowles KH, et al. Clinical and demographic profiles of home care patients with Alzheimer's disease and related dementias: implications for information transfer across care settings. J Appl Gerontol. 2022;41(2):534-544. [FREE Full text] [doi: 10.1177/0733464821999225] [Medline: 33749369]
10. Burgdorf JG, Amjad H, Barrón Y, Ryvicker M. Undocumented dementia diagnosis during skilled home health care: prevalence and associated factors. J Am Geriatr Soc. 2025;73(7):2117-2126. [doi: 10.1111/jgs.19491] [Medline: 40318054]
11. 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. [doi: 10.1016/j.artmed.2023.102624] [Medline: 37673583]

12. Zolnoori M, Barrón Y, Song J, Noble J, Burgdorf J, Ryvicker M, et al. HomeADScreen: developing Alzheimer's disease and related dementia risk identification model in home healthcare. Int J Med Inform. 2023;177:105146. [doi: 10.1016/j.ijmedinf.2023.105146] [Medline: 37454558]

13. Azadmaleki H, Haghbin Y, Rashidi S, Momeni Nezhad MJ, Zolnour A, Zolnoori M. SpeechCARE: dynamic multimodal modeling for cognitive screening in diverse linguistic and speech task contexts. NPJ Digit Med. 2025;8(1):677. [FREE Full text] [doi: 10.1038/s41746-025-02026-x] [Medline: 41249382]

14. Azadmaleki H, Zolnour A, Rashidi S, Noble JM, Hirschberg J, Esmaeili E, et al. TransformerCARE: a novel speech analysis pipeline using transformer-based models and audio augmentation techniques for cognitive impairment detection. Int J Med Inform. 2026;207:106208. [doi: 10.1016/j.ijmedinf.2025.106208] [Medline: 41353845]

15. Jafari Z, Andrew M, Rockwood K. Diagnostic utility of speech-based biomarkers in mild cognitive impairment: a systematic review and meta-analysis. Age Ageing. 2025;54(10):afaf316. [doi: 10.1093/ageing/afaf316] [Medline: 41148189]

16. Zolnoori M, Zolnour A, Vergez S, Sridharan S, Spens I, Topaz M, et al. Beyond electronic health record data: leveraging natural language processing and machine learning to uncover cognitive insights from patient-nurse verbal communications. J Am Med Inform Assoc. 2025;32(2):328-340. [doi: 10.1093/jamia/ocae300] [Medline: 39667364]

17. Luz S, Haider F, de LFS, Fromm D, MacWhinney B. Alzheimer's dementia recognition through spontaneous speech: The ADReSS challenge. 2020. Presented at: Proceedings of the Annual Conference of the International Speech Communication Association. INTERSPEECH; 2020 Oct 25-29:2172-2176; Shanghai, China. [doi: 10.21437/interspeech.2020-2571]

18. Taherinezhad F, Nezhad M, Karimi S. Speech-based cognitive screening: a systematic evaluation of LLM adaptation strategies. arXiv. Preprint posted online on August 18, 2025. [doi: 10.2196/preprints.82608]

19. Zolnour A, Azadmaleki H, Haghbin Y, Taherinezhad F, Nezhad MJM, Rashidi S, et al. LLMCARE: early detection of cognitive impairment via transformer models enhanced by LLM-generated synthetic data. Front Artif Intell. 2025;8:1669896. [FREE Full text] [doi: 10.3389/frai.2025.1669896] [Medline: 41280882]

20. Zolnoori M, Azadmaleki H, Haghbin Y. National Institute on Aging PREPARE challenge: early detection of cognitive impairment using speech-the SpeechCARE solution. arXiv. Preprint posted online on November 11, 2025. [doi: 10.48550/arXiv.2511.08132]

21. Azadmaleki H, Haghbin Y, Rashidi S, Momeni Nezhad MJ, Naserian M, Esmaeili E, 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. [doi: 10.3233/SHTI251249] [Medline: 40776266]

22. Zolnoori M, Vergez S, Kostic Z, Jonnalagadda SR, V McDonald M, Bowles KKH, 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. [FREE Full text] [doi: 10.2196/35325] [Medline: 35544296]

23. Topaz M, Barrón Y, Song J, Onorato N, Sockolow P, Zolnoori M, et al. Risk of rehospitalization or emergency department visit is significantly higher for patients who receive their first home health care nursing visit later than 2 days after hospital discharge. J Am Med Dir Assoc. 2022;23(10):1642-1647. [doi: 10.1016/j.jamda.2022.07.001] [Medline: 35931136]

24. Shankar R, Bundele A, Mukhopadhyay A. Natural language processing of electronic health records for early detection of cognitive decline: a systematic review. NPJ Digit Med. 2025;8(1):133. [FREE Full text] [doi: 10.1038/s41746-025-01527-z] [Medline: 40025194]

25. 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. 2025;57(6):1003-1011. [doi: 10.1111/jnu.70039] [Medline: 40785044]

26. Chen YW, Ho W, Vergez SM. Hearing health in home healthcare: leveraging LLMs for Illness scoring and ALMs for vocal biomarker extraction. arXiv. Preprint posted online on October 20, 2025

27. Zhang Z, Momeni Nezhad MJ, Gupta P, Zolnour A, Azadmaleki H, Topaz M, et al. Enhancing AI for citation screening in literature reviews: improving accuracy with ensemble models. Int J Med Inform. 2025;203:106035. [doi: 10.1016/j.ijmedinf.2025.106035] [Medline: 40609462]

28. Hosseini S, Momeni Nezhad MJ, Hosseini M, Zolnoori M. Optimizing entity recognition in psychiatric treatment data with large language models. Stud Health Technol Inform. 2025;329:784-788. [doi: 10.3233/SHTI250947] [Medline: 40775965]

29. Zhang Z, Nezhad M, Hosseini S, Zolnour A, Zonour Z, Hosseini SM, et al. A scoping review of large language model applications in healthcare. Stud Health Technol Inform. 2025;329:1966-1967. [doi: 10.3233/SHTI251302] [Medline: 40776319]

30. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc. 2005;53(4):695-699. [doi: 10.1111/j.1532-5415.2005.53221.x] [Medline: 15817019]

31. Berg L. Clinical dementia rating. Br J Psychiatry. 2018;145(3):339-339. [doi: 10.1192/s0007125000118082]

32. Saramonic wireless microphone. Amazon. URL: https://tinyurl.com/3cwrvzpz [accessed 2025-12-17]

33. Kinatukara S, Rosati RJ, Huang L. Assessment of OASIS reliability and validity using several methodological approaches. Home Health Care Serv Q. 2005;24(3):23-38. [doi: 10.1300/J027v24n03_02] [Medline: 16203688]

34. Tullai-McGuinness S, Madigan EA, Fortinsky RH. Validity testing the Outcomes and Assessment Information Set (OASIS). Home Health Care Serv Q. 2009;28(1):45-57. [FREE Full text] [doi: 10.1080/01621420802716206] [Medline: 19266370]

35. RX 8: great audio starts with RX. IZOTOPE. URL: https://www.izotope.com/en/products/rx.html [accessed 2026-01-13]

36. Zolnoori M, Vergez S, Sridharan S, Zolnour A, Bowles K, Kostic Z, 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] [Medline: 37478477]

37. Eyben F, Schuller B. The Munich open-source large-scale multimedia feature extractor. ACM SIGMultimedia Rec. 2015;6(4):4-13. [doi: 10.1145/2729095.2729097]

38. Praat Vocal Toolkit. URL: http://www.praatvocaltoolkit.com/ [accessed 2021-04-19]

39. Eyben F, Scherer KR, Schuller BW, Sundberg J, Andre E, Busso C, et al. The geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. IEEE Trans Affective Comput. 2016;7(2):190-202. [doi: 10.1109/taffc.2015.2457417]

40. Pennebaker J, Boyd R, Jordan K, Blackburn K. The development and psychometric properties of LIWC2015. University of Texas at Austin. 2015. URL: https://www.liwc.app/ [accessed 2026-01-21]

41. Bird S, Klein E, Loper E. Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. Santa Rosa, CA. O'Reilly Media Inc; 2009.

42. Liu Y, Ott M, Goyal N. Roberta: a robustly optimized bert pretraining approach. arXiv. Preprint posted online on July 26, 2019. [doi: 10.48550/arXiv.1907.11692]

43. Lanzi AM, Saylor AK, Fromm D, Liu H, MacWhinney B, Cohen ML. DementiaBank: theoretical rationale, protocol, and illustrative analyses. Am J Speech Lang Pathol. 2023;32(2):426-438. [FREE Full text] [doi: 10.1044/2022_AJSLP-22-00281] [Medline: 36791255]

44. 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. Behav Neurol. 2020;2020:4683573. [FREE Full text] [doi: 10.1155/2020/4683573] [Medline: 32351632]

45. 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. [FREE Full text] [doi: 10.1371/journal.pone.0236009] [Medline: 32658934]

46. Huet K, Delvaux V, Piccaluga M, Roland V, Harmegnies B. Inter-syllabic interval as an indicator of fluency in Parkinsonian French speech. 2017. Presented at: 11th International Seminar on Speech Production; 2017 October 16-19; Tianjin, China. [doi: 10.21437/interspeech.2016-349]

47. 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. [FREE Full text] [doi: 10.3389/fpsyg.2021.620251] [Medline: 33833713]

48. Calzà L, Gagliardi G, Rossini Favretti R, Tamburini F. Linguistic features and automatic classifiers for identifying mild cognitive impairment and dementia. Comput Speech Lang. 2021;65:101113. [doi: 10.1016/j.csl.2020.101113]

49. Dahmani H, Selouani S, O'shaughnessy D, Chetouani M, Doghmane N. Assessment of dysarthric speech through rhythm metrics. J King Saud Univ Comput Inf Sci. 2013;25(1):43-49. [doi: 10.1016/j.jksuci.2012.05.005]

50. Boersma P. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. 1993. Presented at: Proceedings of the Institute of Phonetic Sciences; 2026 Jan 13; Durham, North Carolina.

51. Viegas F, Viegas D, Guimarães GS, Souza MMGD, Luiz RR, Simões-Zenari M, et al. Comparison of fundamental frequency and formants frequency measurements in two speech tasks. Revista CEFAC. 2019;21(6). [doi: 10.1590/1982-0216/201921612819]

52. Kong Y, Mullangi A, Marozeau J, Epstein M. Temporal and spectral cues for musical timbre perception in electric hearing. J Speech Lang Hear Res. 2011;54(3):981-994. [FREE Full text] [doi: 10.1044/1092-4388(2010/10-0196)] [Medline: 21060140]

53. Tjaden K, Sussman JE, Liu G, Wilding G. Long-term average spectral (LTAS) measures of dysarthria and their relationship to perceived severity. J Med Speech Lang Pathol. 2010;18(4):125-132. [FREE Full text] [Medline: 29422763]

54. On C, Pandiyan P, Yaacob S, Saudi A. Mel-frequency cepstral coefficient analysis in speech recognition. 2006. Presented at: International Conference on Computing & Informatics; 2006 June 6-8:1-5; Kuala Lumpur. [doi: 10.1109/icoci.2006.5276486]

55. Teixeira JP, Oliveira C, Lopes C. Vocal acoustic analysis–jitter, shimmer and HNR parameters. Procedia Technology. 2013;9:1112-1122. [doi: 10.1016/j.protcy.2013.12.124]

56. Fraile R, Godino-Llorente JI. Cepstral peak prominence: a comprehensive analysis. Biomed Signal Process Control. Nov 2014;14:42-54. [doi: 10.1016/j.bspc.2014.07.001]

57. Simonyan K, Tovar-Moll F, Ostuni J, Hallett M, Kalasinsky VF, Lewin-Smith MR, et al. Focal white matter changes in spasmodic dysphonia: a combined diffusion tensor imaging and neuropathological study. Brain. 2008;131(Pt 2):447-459. [FREE Full text] [doi: 10.1093/brain/awm303] [Medline: 18083751]

58. Maryn Y, De Bodt M, Roy N. The Acoustic Voice Quality Index: toward improved treatment outcomes assessment in voice disorders. J Commun Disord. 2010;43(3):161-174. [doi: 10.1016/j.jcomdis.2009.12.004] [Medline: 20080243]

59. Tamarit L, Goudbeek M, Scherer K. Spectral slope measurements in emotionally expressive speech. 2008. Presented at: Proceedings of Speech Analysis and Processing for Knowledge Discovery; 2008 June 4-6:169-183; Aalborg, Denmark.

60. Deutsch D, Henthorn T, Dolson M. Absolute pitch, speech, and tone languageome experiments and a proposed framework. Music Percept. 2004;21(3):339-356. [doi: 10.1525/mp.2004.21.3.339]

61. Gramming P, Sundberg J, Ternström S, Leanderson R, Perkins WH. Relationship between changes in voice pitch and loudness. J Voice. 1988;2(2):118-126. [doi: 10.1016/s0892-1997(88)80067-5]

62. Narasimhan SV, Vishal K. Spectral measures of hoarseness in persons with hyperfunctional voice disorder. J Voice. 2017;31(1):57-61. [doi: 10.1016/j.jvoice.2016.03.005]

63. Fergadiotis G, Wright HH, Green SB. Psychometric evaluation of lexical diversity indices: assessing length effects. J Speech Lang Hear Res. 2015;58(3):840-852. [FREE Full text] [doi: 10.1044/2015_JSLHR-L-14-0280] [Medline: 25766139]

64. Sanborn V, Ostrand R, Ciesla J, Gunstad J. Automated assessment of speech production and prediction of MCI in older adults. Appl Neuropsychol Adult. 2020;29(5):1-8. [FREE Full text] [doi: 10.1080/23279095.2020.1864733] [Medline: 33377800]

65. 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. 2020;2:567158. [FREE Full text] [doi: 10.3389/fdgth.2020.567158] [Medline: 34713039]

66. Cheung H, Kemper S. Competing complexity metrics and adults' production of complex sentences. Appl Psycholinguist. 1992;13(1):53e76. [FREE Full text] [doi: 10.1371/journal.pbio.3002968] [Medline: 39836653]

67. Lakretz Y, Dehaene S, King J. What limits our capacity to process nested long-range dependencies in sentence comprehension? Entropy (Basel). 2020;22(4):446. [FREE Full text] [doi: 10.3390/e22040446] [Medline: 33286220]

68. Yeung C, Lee J. Automatic detection of sentence fragments. 2015. Presented at: The 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing; 2015 July 26-31:599-603; Beijing, China. [doi: 10.3115/v1/p15-2099]

69. López-de-Ipiña K, Martinez-de-Lizarduy U, Calvo PM, Beitia B, García-Melero J, Fernández E, et al. On the analysis of speech and disfluencies for automatic detection of mild cognitive impairment. Neural Comput Applic. 2018;32(20):15761-15769. [doi: 10.1007/s00521-018-3494-1]

70. Tóth L, Gosztolya G, Vincze V. Automatic detection of mild cognitive impairment from spontaneous speech using ASR. 2015. Presented at: Sixteenth Annual Conference of the International Speech Communication Association; 2015 September 6-10; Dresden, Germany. [doi: 10.21437/interspeech.2015-568]

71. Xu Z, Vergez S, Esmaeili E, Zolnour A, Briggs KA, Scroggins JK, 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. [doi: 10.3233/SHTI251273] [Medline: 40776290]

72. Zolnoori M, Vergez S, Xu Z, Esmaeili E, Zolnour A, Anne Briggs K, 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. [doi: 10.1093/jamiaopen/ooae130] [Medline: 39659993]

73. Ferson S, O'Rawe J, Antonenko A, Siegrist J, Mickley J, Luhmann CC, et al. Natural language of uncertainty: numeric hedge words. Int J Approx Reason. 2015;57:19-39. [doi: 10.1016/j.ijar.2014.11.003]

74. 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 J Audio Speech Music Process. 2015;2015(1):9. [doi: 10.1186/s13636-015-0052-y]

75. Topaz M. NimbleMiner: a novel multi-lingual text mining application. Stud Health Technol Inform. 2019;264:1608-1609. [doi: 10.3233/SHTI190558] [Medline: 31438255]

76. 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. [FREE Full text] [doi: 10.1016/j.pec.2007.10.019] [Medline: 18093788]

77. Drew P, Chatwin J, Collins S. Conversation analysis: a method for research into interactions between patients and health-care professionals. Health Expect. 2001;4(1):58-70. [FREE Full text] [doi: 10.1046/j.1369-6513.2001.00125.x] [Medline: 11286600]

78. Eyben F, Wöllmer M, Schuller B. Opensmile: the munich versatile and fast open-source audio feature extracto. 2010. Presented at: Proceedings of the 18th ACM International Conference on Multimedia; 2010 oct 25-29:1459-1462; Firenze, Italy. [doi: 10.1145/1873951.1874246]

79. Bahgat M, Wilson S, Magdy W. Classifying online slang terms into LIWC categories. 2022. Presented at: 14th ACM Web Science Conference; 2022 Jun 26-29:422-432; Barcelona, Spain. [doi: 10.1145/3501247.3531572]

80. Duffy JR, Josephs KA. The diagnosis and understanding of apraxia of speech: why including neurodegenerative etiologies may be important. J Speech Lang Hear Res. 2012;55(5):S1518-S1522. [FREE Full text] [doi: 10.1044/1092-4388(2012/11-0309)] [Medline: 23033445]

81. Ward M, Cecato JF, Aprahamian I, Martinelli JE. Assessment for apraxia in mild cognitive impairment and Alzheimer's dise. Dement Neuropsychol. 2015;9(1):71-75. [FREE Full text] [doi: 10.1590/S1980-57642015DN91000011] [Medline: 29213944]

82.   Nagumo R, Zhang Y, Ogawa Y, Hosokawa M, Abe K, Ukeda T, et al. Automatic detection of cognitive impairments through
       acoustic analysis of speech. Curr Alzheimer Res. 2020;17(1):60-68. [doi: 10.2174/1567205017666200213094513] [Medline:
       32053074]

83.   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? Front Aging Neurosci. 2018;10:369.
       [FREE Full text] [doi: 10.3389/fnagi.2018.00369] [Medline: 30483116]

84.   Themistocleous C, Eckerström M, Kokkinakis D. Identification of mild cognitive impairment from speech in swedish using
       deep sequential neural networks. Front Neurol. 2018;9:975. [FREE Full text] [doi: 10.3389/fneur.2018.00975] [Medline:
       30498472]

85.   Bidelman GM, Lowther JE, Tak SH, Alain C. Mild cognitive impairment is characterized by deficient brainstem and cortical
       representations of speech. J Neurosci. 2017;37(13):3610-3620. [FREE Full text] [doi: 10.1523/JNEUROSCI.3700-16.2017]
       [Medline: 28270574]

86.   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. [doi: 10.1109/access.2018.2856478]

87.   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. [doi:
       10.1371/journal.pone.0217388]

88.   Mirzaei S, El Yacoubi M, Garcia-Salicetti S, Boudy J, Kahindo C, Cristancho-Lacroix V, 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]

89.   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. Alzheimers Dement (Amst). 2015;1(1):112-124. [FREE Full text]
       [doi: 10.1016/j.dadm.2014.11.012] [Medline: 27239498]

90.   Ning L, Luo K. Using text and acoustic features to diagnose mild cognitive impairment and Alzheimer's disease. Research
       Square. Preprint posted online on October 21, 2020. [doi: 10.21203/rs.3.rs-92702/v1]

91.   Han K-H, Zaytseva Y, Bao Y, Pöppel E, Chung SY, Kim JW, et al. Impairment of vocal expression of negative emotions
       in patients with Alzheimer's disease. Front Aging Neurosci. 2014;6:101. [FREE Full text] [doi: 10.3389/fnagi.2014.00101]
       [Medline: 24904413]

92.   Cadieux NL, Greve KW. Emotion processing in Alzheimer's disease. J Int Neuropsychol Soc. 1997;3(5):411-419. [Medline:
       9322399]

93.   Kamiloğlu RG, Fischer AH, Sauter DA. Good vibrations: a review of vocal expressions of positive emotions. Psychon Bull
       Rev. 2020;27(2):237-265. [FREE Full text] [doi: 10.3758/s13423-019-01701-x] [Medline: 31898261]

94.   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. Front Psychol. 2013;4:292. [FREE Full text] [doi: 10.3389/fpsyg.2013.00292] [Medline:
       23750144]

95.   Spazzapan EA, de Castro Marino VC, Cardoso VM, Berti LC, Fabron EMG. Acoustic characteristics of voice in different
       cycles of life: an integrative literature review. Revista CEFAC. 2019;21(3). [doi: 10.1590/1982-0216/201921315018]

96.   Grabowski K, Rynkiewicz A, Lassalle A, Baron-Cohen S, Schuller B, Cummins N, et al. Emotional expression in psychiatric
       conditions: new technology for clinicians. Psychiatry Clin Neurosci. 2019;73(2):50-62. [FREE Full text] [doi:
       10.1111/pcn.12799] [Medline: 30565801]

97.   Asgari M, Kaye J, Dodge H. Predicting mild cognitive impairment from spontaneous spoken utterances. Alzheimers Dement
       (NY). 2017;3(2):219-228. [FREE Full text] [doi: 10.1016/j.trci.2017.01.006] [Medline: 29067328]

98.   Shen J, Rudzicz F. Detecting anxiety through reddit. 2017. Presented at: Proceedings of the Fourth Workshop on
       Computational Linguistics and Clinical Psychology — From Linguistic Signal to Clinical Reality; 2017 August 01:58-65;
       Vancouver, BC.

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. J Speech Lang Hear Res. 2020;63(5):1416-1429. [doi: 10.1044/2020_JSLHR-19-00335]
       [Medline: 32402217]

100.   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. [FREE Full text] [doi: 10.1371/journal.pone.0155195] [Medline: 27176919]

101.   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. [FREE Full text]
       [doi: 10.1080/13803395.2018.1446513] [Medline: 29669461]

102.   Ostrand R, Gunstad J. Using automatic assessment of speech production to predict current and future cognitive function
       in older adults. J Geriatr Psychiatry Neurol. 2021;34(5):357-369. [FREE Full text] [doi: 10.1177/0891988720933358]
       [Medline: 32723128]

103.   Davy W, Travis J, Laura W, Dona L, Graciela G. Towards automatic detection of abnormal cognitive decline and dementia
       through linguistic analysis of writing samples. 2016. Presented at: 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 June 12-17:1198-1207; San Diego, California. [doi: 10.18653/v1/n16-1143]

104. Gosztolya G, Tóth L, Grósz T. Detecting mild cognitive impairment from spontaneous speech by correlation-based phonetic feature selection. 2016. Presented at: 17th Annual Conference of the International Speech Communication Association; 2016, September 8-12; San Francisco. [doi: 10.21437/interspeech.2016-384]

105. Johnson M, Lin F. Communication difficulty and relevant interventions in mild cognitive impairment: implications for neuroplasticity. Top Geriatr Rehabil. 2014;30(1):18-34. [FREE Full text] [doi: 10.1097/TGR.0000000000000001] [Medline: 25356002]

106. Pan C-W, Wang X, Ma Q, Sun H, Xu Y, Wang P. Cognitive dysfunction and health-related quality of life among older Chinese. Sci Rep. 2015;5:17301. [FREE Full text] [doi: 10.1038/srep17301] [Medline: 26601612]

107. Ilan S, Carmel S. Patient communication pattern scale: psychometric characteristics. Health Expect. 2016;19(4):842-853. [FREE Full text] [doi: 10.1111/hex.12381] [Medline: 26172267]

108. Dodge HH, Mattek N, Gregor M, Bowman M, Seelye A, Ybarra O, et al. Social markers of mild cognitive impairment: proportion of word counts in free conversational speech. Curr Alzheimer Res. 2015;12(6):513-519. [FREE Full text] [doi: 10.2174/1567205012666150530201917] [Medline: 26027814]

109. Seinen TM, Kors JA, van Mulligen EM, Rijnbeek PR. Using structured codes and free-text notes to measure information complementarity in electronic health records: feasibility and validation study. J Med Internet Res. 2025;27:e66910. [FREE Full text] [doi: 10.2196/66910] [Medline: 39946687]

110. Song J, Zolnoori M, Scharp D, Vergez S, McDonald MV, Sridharan S, et al. Do nurses document all discussions of patient problems and nursing interventions in the electronic health record? A pilot study in home healthcare. JAMIA Open. 2022;5(2):ooac034. [FREE Full text] [doi: 10.1093/jamiaopen/ooac034] [Medline: 35663115]

111. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res. 2004;32(Database issue):D267-D270. [FREE Full text] [doi: 10.1093/nar/gkh061] [Medline: 14681409]

112. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb). 2012;22(3):276-282. [FREE Full text] [Medline: 23092060]

113. Zhang S, Dong L, Li X, Zhang S, Sun X, Wang S, et al. Instruction tuning for large language models: a survey. arXiv. Preprint posted online on August 21, 2023. [doi: 10.48550/arXiv.2308.10792]

114. Hu EJ, Shen Y, Wallis P, Allen-Zhu Z, Li Y, Wang S, et al. Lora: low-rank adaptation of large language models. arXiv. Preprint posted online on June 17, 2021. [doi: 10.48550/arXiv.2106.09685]

115. Li Y, Yu Y, Liang C, He P, Karampatziakis N, Chen W, et al. LoftQ: lora-fine-tuning-aware quantization for large language models. arXiv. Preprint posted online on October 12, 2023

116. 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. [FREE Full text] [doi: 10.2196/18366] [Medline: 33118958]

117. 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]

118. Murty MN, Raghava R. Kernel-Based SVM. In: Support Vector Machines and Perceptrons. Cham. Springer; 2016:57-67.

119. Zhou ZH, Tang W. Selective ensemble of decision trees. In: Wang G, Liu Q, Yao Y, Skowron A, editors. International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing. Berlin. Springer; 2003:476-483.

120. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. Mach Learn. 2006;63(1):3-42. [doi: 10.1007/s10994-006-6226-1]

121. Chen T, Guestrin C. Xgboost: a scalable tree boosting system. 2016. Presented at: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016 August 13:785-794; California, San Francisco, USA.

122. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. Lightgbm: A highly efficient gradient boosting decision tree. Adv Neural Inf Process Syst. 2017;30:3146-3154.

Abbreviations

AUC-ROC: area under the receiver operating characteristic curve

AWS: Amazon Web Services

CDR: Clinical Dementia Rating

EHR: electronic health record

GeMAPS: Geneva Minimalistic Acoustic Parameter Set

HHC: home health care

ICD-10: International Statistical Classification of Diseases, Tenth Revision

LIWC: Linguistic Inquiry and Word Count

LLM: large language model

MCI: mild cognitive impairment

MCI-ED: mild cognitive impairment and early dementia

MoCA: Montreal Cognitive Assessment

NLP: natural language processing

NLTK: Natural Language Toolkit
OASIS: Outcome and Assessment Information Set
SVM: support vector machine
UMLS: Unified Medical Language System

XSL•FO
RenderX