Towards Scalable Remote Assessment of Mild Cognitive Impairment Via Multimodal Dialog

Oliver Roesler[1], Jackson Liscombe[1,2], Michael Neumann[1], Hardik Kothare[1,2], Abhishek Hosamath[1], Lakshmi Arbatti[1], Doug Habberstad[1,2], Christiane Suendermann-Oeft[1,2], Meredith Bartlett[1], Cathy Zhang[1], Nikhil Sukhdev[1], Kolja Wilms[1], Anusha Badathala[2,4], Sandrine Istas[5], Steve Ruhmel[3], Bryan Hansen[3], Madeline Hannan[3], David Henley[3], Arthur Wallace[2,4], Ira Shoulson[1], David Suendermann-Oeft[1], Vikram Ramanarayanan[1,2,4]

[1]Modality.AI, Inc., San Francisco, CA, USA. [2]San Francisco Veterans Affairs Health Care System, San Francisco, CA, USA. [3]Janssen Research and Development, LLC., NJ, USA. [4]University of California, San Francisco, San Francisco, CA, USA. [5]ICON Strategic Solutions, ICON plc., Belgium

v@modality.ai

Abstract

Early assessment of mild cognitive impairment (MCI) has the potential to expedite interventions and slow disease progress for people at risk of developing dementia. We investigate the feasibility of administering remote assessments of speech, orofacial and cognitive function to an elderly population with MCI via a cloud-based conversational remote monitoring platform, and the utility of automatically extracted multimodal biomarkers and self-reported problems in identifying MCI patients. We analyzed data from 90 MCI patients and 91 controls who each completed two assessments. 90% of participants reported excellent engagement and liked their overall user experience. Furthermore, combining multiple facial, speech and cognitive markers performed best at distinguishing MCI patients from controls with an AUC of 0.75 using a support vector machine classifier. Finally, we found that MCI patients reported significantly more problems related to memory, falls, anxiety and speech than controls.

Index Terms: multimodal dialog system, remote patient monitoring, mild cognitive impairment

1. Introduction

Mild cognitive impairment (MCI) describes cognitive decline that is stronger than the decline expected due to normal aging for people with a similar age and educational background but does not count as dementia because it does not significantly interfere with activities of daily living [1]. About 10-20% of adults who are at least 65 years old have MCI. Men have a higher risk than women to develop MCI and the risk increases with age [2], while people with MCI have a greater risk of developing dementia than people without MCI, although not everyone with MCI will develop dementia [3, 4]. Therefore, identifying people with MCI has the potential to allow for early pharmaceutical interventions before strong damage to the central nervous system has occurred [5]. However, it is non-trivial to detect MCI even for experts – nearly 50% of MCI patients are never diagnosed with MCI [6].

Several studies have demonstrated the utility of speech and video signals for the assessment of MCI in particular, and dementias and other neurological conditions more broadly [7, 8, 9, 10, 11]. For example, Vincze et al. [12] extracted a variety of linguistic features from spontaneous speech transcripts of 48 MCI patients and 36 healthy controls and achieved an accuracy of 75% when training a support vector machine (SVM) for binary classification using leave-one-out cross-validation (LOOCV) with only the features that showed a statistically significant different between cohorts. Similar results were achieved by Asgari et al. [13] who transcribed unstructured conversations of 14 MCI patients and 27 healthy controls to extract several linguistic features that were provided as input to a SVM, which achieved an accuracy of 83% and a maximum AUC of 0.80. In another study, Fraser et al. [14] collected data from 26 MCI patients and 29 healthy controls for three tasks and extracted a variety of features, i.e. speech and language features for a picture description task, eye-tracking and comprehension features for a silent reading task, and eye-tracking, comprehension, and speech features for an aloud reading task. Afterwards, they evaluated different ways of combining the different modalities and used both logistic regression and SVM classifiers for binary classification experiments using LOOCV. For both classifiers the maximum AUC was 0.88.

While the classification results in the aforementioned studies and other similar ones are promising, there are several limitations. First, it is not clear how generalizable the results are due to the relatively small sample sizes, which is a common limitation due to the effort and cost of recruiting and assessing participants, because several studies have shown that accuracies reported for studies with small sample sizes are often overoptimistic [15, 16]. Secondly, many of these studies typically analyze a single modality, i.e., text or speech, in isolation, as opposed to combining information from multiple modalities. Thirdly, many of these studies were either partly or wholly performed in-lab or in-clinic, with data collection technologies not built for scale. Finally, in addition to objective measures of patient behavior captured by such conversational remote patient monitoring technologies, what patients report about their illness is also of critical importance, but has traditionally been captured using categorical scales that are rated by clinicians in research settings. However, recent research has demonstrated the efficacy of analyzing open-ended self-reported responses from patients to questions about what bothers them about their disease and how it affects their daily functioning [17].

To address these limitations, we propose a multimodal dialog platform for remote assessment of MCI that employs a virtual agent to guide a sample of 181 participants through a number of tasks while extracting a variety of speech, text, facial and cognitive features. While there are many telehealth solutions for remote patient assessment and monitoring, to our knowledge, this is the first work that integrates both (i) the measurement of objective biomarkers of MCI from multiple modalities and (ii) unfiltered patient verbatim replies about their bothersome problems and functional consequences into one comprehensive solution at scale. This paper specifically aims to answer the following research questions:

1. Is it feasible to remote administer assessments of speech and cognitive function at scale through a multimodal dialog platform for an elderly population with MCI?
2. How informative and reliable are the extracted features at distinguishing MCI patients from healthy controls? Moreover, what additional insights do analyses of MCI patients' most bothersome problems relative to healthy controls reveal about the disease?

To this end, we collected 362 remote assessments from 90 people with MCI and 91 healthy controls and analyzed both the self-reported experience of the participants as well as the utility and reliability of the extracted features.

2. Data

We recruited 200 participants (100 people with MCI and 100 healthy controls) via the U.S. Department of Veterans Affairs[1] between November 2023 and January 2024 to complete two assessments (one week apart) administered through the Modality platform, a HIPAA compliant cloud-based multimodal dialog platform [18, 19]. To qualify as participants, people needed to be at least 55 years old, able to consent and e-sign, have a valid phone number and email, able to read and speak in English, and have access to a smartphone, tablet, or PC with internet connection and webcam. Additionally, people were not allowed to participate, if they had been diagnosed with dementia, had cognitive impairment due to cerebrovascular disease, head trauma, chronic or active abuse of alcohol, opioid, or methamphetamines, Parkinson's disease, schizophrenia, bipolar disease, or major depressive disorder, or used benzodiazepines, non-BZD receptor modulator sleeping medications, drugs for the treatment of Parkinson's disease such as levodopa, or antipsychotics. Finally, for the MCI cohort participants needed at least two MCI diagnoses (ICD-10[2]).

During each assessment participants were guided through 23 structured exercises designed to elicit speech, facial, and cognitive behaviors, including vowel phonations, automatic speech (counting up from one), read speech (SIT and bamboo passage), spontaneous speech (picture description and open ended questions), cognitive recall (immediate and delayed), digit span (forward and backward), three step task, and categorical fluency. Additionally, during the first assessment at the end of the interactive part, participants were asked by the virtual agent whether they have a problem related to their general health or personal well being, to describe in their own words the problem and how it affects their daily functioning, and to describe what makes the problem better and worse. Participants were asked to report up to five problems related to their general health as well as up to five problems related to their personal well being. Finally, at the beginning of the first assessment participants were presented with a demographics survey before the interactive part of the assessment and at the end of each assessment participants were asked to complete a user experience survey. The study was approved by the Institutional Review Board of the University of California, San Francisco.

[1]https://www.usa.gov/agencies/u-s-department-of-veterans-affairs
[2]https://www.cdc.gov/nchs/icd/icd10.htm

Table 1: *Participant demographics. Age is presented as mean (standard deviation).*

Cohort	# Participants	Age (years)
MCI	90 (9F / 81M)	71.08 (9.10)
Controls	91 (9F / 82M)	71.30 (8.59)

In the end, 181 participants completed both assessments, while 19 participants withdrew from the study for a variety of reasons. Table 1 provides an overview of the demographics of the 181 participants. While the cohorts have been age-matched, about 90% of participants were male reflecting the fact that only about 10% of US veterans are female[3]. Of the healthy controls, about 92% identified themselves as white, 5% as black, and 3% as others, while for MCI patients 80% identified themselves as white, 16% as black, and 4% as others. The education level was similar for both cohorts with 44% of the MCI patients and 35% of the controls having an advanced degree, 47% of the MCI patients and 60% of the controls having an undergraduate degree, and 9% of the MCI patients 5% of the controls having a high school degree or GED[4].

3. Feature Extraction

The multimodal dialog platform automatically extracts a variety of speech (e.g. speaking rate), facial (e.g. range and speed of movement of the lips), and text (e.g. noun rate) features in near-real-time during interactive assessments. Speech features are extracted using Praat [20] and Kaldi [21]. Facial features are computed using facial landmarks extracted with MediaPipe Face Detection [22] and MediaPipe Face Mesh [23], and normalized through the inter-caruncular distance to reduce the variability within and across assessments due to camera and head movement. Text features were computed using SpaCy 3.7.2 [5] based on automatic transcriptions obtained through Amazon Transcribe[6]. Cognitive features (e.g. number of words recalled) were manually extracted by human annotators after the data collection. For the word recall tasks, the score is expressed as the percentage of correct words (ignoring order). For the digit span tasks, a score of 2 was given if all digits were repeated in the correct order, a score of 1 if all digits were repeated but in a different order, and a score of 0 otherwise. Additionally, for all cognitive tasks, the end of the system's prompt and the beginning as well as the end of the participant's response was manually determined and used to compute response latency and response duration features.

4. Feasibility Analysis

To evaluate the feasibility of conducting a remote assessment through a multimodal dialog platform, participants were asked to rate different aspects of the interaction, like how engaged they felt during the interaction or how intelligible the virtual agent was, on a 5-point Likert scale ranging from "very unsatisfactory" to "very satisfactory" after the interactive part of the session. Figure 1 shows that the majority of participants rated the relatability, understandability, and intelligibility of the system's voice as very high, and 98.9% of the participants indicated that they were never or only rarely interrupted by the system. Additionally, most of the participants responded that the overall

[3]https://www.va.gov/vetdata/veteran_population.asp
[4]https://ged.com/
[5]https://spacy.io/
[6]https://aws.amazon.com/transcribe/

Table 2: *Overview of the extracted features across modalities.*

	Domain	Features
Speech	Energy	shimmer (%), intensity (dB), signal-to-noise ratio (dB)
	Timing	speaking and articulation duration (sec.), articulation and speaking rate (WPM), percent pause time (PPT, %), canonical timing agreement (CTA, %)
	Voice quality	cepstral peak prominence (CPP, dB), harmonics-to-noise ratio (HNR, dB)
	Frequency	mean, max., min. fundamental frequency F0 (Hz), first three formants F1, F2, F3 (Hz), slope of 2nd formant (Hz/sec.), jitter (%)
Facial	Mouth measurements	lip aperture/opening, lip width, mouth surface area, mean symmetry ratio between left and right half of the mouth
	Movement	velocity, acceleration, jerk, and speed of lower lip and jaw center
	Eyes	number of eye blinks per sec., eye opening, vertical displacement of eyebrows
Text	Lexico-semantic	word count, percentage of content words, noun rate, verb rate, pronoun rate, noun-to-verb ratio, noun-to-pronoun ratio, closed class word ratio, idea density
	Self-reported problems	reported symptoms, reported problem domains
Cognitive	Scores	percentage of correct words (immediate and delayed word recall), digit span forward/backward score (ranges from 0 to 2)
	Timing	response latency (sec.), response duration (sec.)

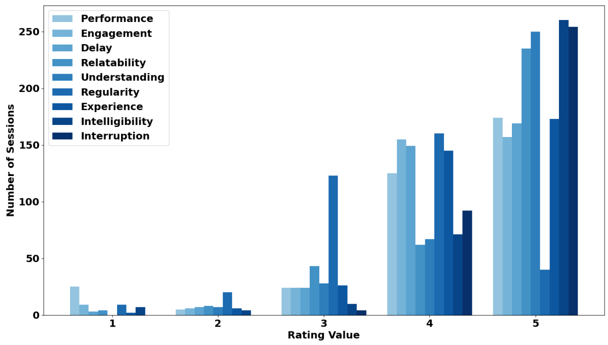

Figure 1: *Bar chart illustrating the results of the UX survey. The x-axis represents ratings on a Likert scale from 1:Very Unsatisfactory to 5:Very Satisfactory*

system's performance, their overall experience with the system, and the delay in response from the system were either satisfactory or very satisfactory. Furthermore, most participants felt either engaged or even highly engaged while interacting with the system. Only when participants were asked how regularly they would use an app with the virtual guide, only 11.4% selected the best rating stating that they would use it all the time, while most participants replied that they would use it often or sometimes, which is not surprising considering that the system was not intended to be used all the time. Overall the results highlight that it is feasible to conduct remote assessments through a multimodal dialog platform for an elderly population with and without MCI, although it is important to remember that only people who have an email address and an electronic device with internet connection and webcam were allowed to participate.

5. Clinical Validation

To determine whether the features extracted from the collected data are clinically valid, non-parametric Kruskal-Wallis tests were performed for each individual feature to determine which of them show a statistically significant difference ($\alpha = 0.01$) between cohorts. Additionally, Pearson correlations were computed between features of participants' subsequent sessions to assess the test-retest reliability of the features. Figure 2 shows the effect sizes in terms of Glass' Delta for the 13 resulting features together with their test-retest reliability. The figure shows that facial features show the strongest signal with mostly acceptable or good reliability. More specifically, features related to the average mouth surface area and lip aperture are higher for MCI patients, while features related to the vertical eyebrow position and eye opening are lower for MCI patients. For picture description the minimum F0 was lower and for the delayed word recall task intensity was lower, however, the latter feature was not reliable. Finally, healthy controls achieved higher delayed recall scores than MCI patients, although this feature was not reliable.

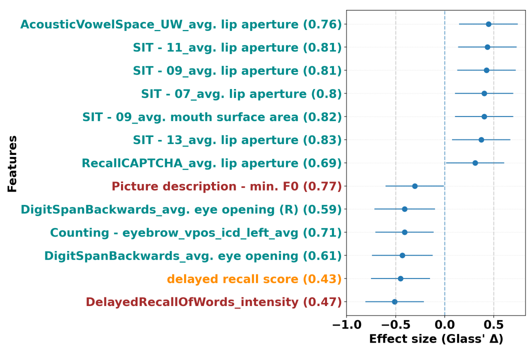

Figure 2: *Effect sizes and test-retest reliabilities (reported in parentheses) of speech, facial, and cognitive features that show statistically significant differences between MCI patients and healthy controls at $\alpha = 0.01$. Positive effect sizes indicate that feature values for patients are greater than for healthy controls.*

To investigate how well the features can discriminate the cohorts, several classifiers available in scikit-learn 1.3.2 [24], i.e. Logistic Regression (*LogisticRegression*), Random Forest (*RandomForestClassifier*), Multilayer Perceptron (*MLPClassifier*), and Support Vector Machine (*SVC*), were employed for binary classification experiments using 5-fold cross-validation. Since some features had missing values due to incomplete dialogs or data transmission errors and most machine learning algorithms cannot handle missing feature values, we applied a two step process to remove missing values: (1) we removed features with more than 5% missing values in the data, and (2) we removed the remaining missing data by taking out affected participant sessions. Afterwards, the following feature sets were provided as input to all classifiers: (i) speech only, (ii) facial only, (iii) text only, (iv) cognitive only, (v) all features combined, and (vi) only features that showed a statistically significant difference between cohorts (Figure 2). The best result was obtained using logistic regression or support vector machine classifiers for feature set (vi) with a mean AUC of 0.75 (Figure 3).

Table 3 illustrates that the selected feature set had the strongest influence on the achieved classification performance, while the influence of the selected classifier was small in comparison. When only a single modality was used, text features led to the the best results with an AUC of 0.62 with logistic regression and MLP classifiers. Providing all features as input to the classifiers without feature selection did not improve the performance, while using only the features that showed a statistically significant difference between cohorts ($p < 0.01$) led to the best AUC of 0.75.

To evaluate the utility of asking participants to report in

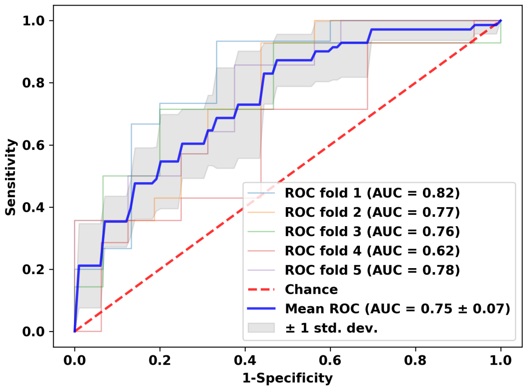

Figure 3: *ROC curves showing the results of binary classification with 5-fold cross-validation when using the 13 features that showed statistically significant differences between controls and MCI patients (see Fig. 2) as input to a support vector machine.*

Table 3: *Classification performance as measured by area under the ROC curve (AUC) across multiple classifiers (LR: Logistic Regression; RF: Random Forests; MLP: Multi-layer Perceptron; SVM: Support Vector Machine) and feature sets.*

		Classifier			
		LR	RF	MLP	SVM
Feature Set	speech only	0.52	0.53	0.55	0.53
	facial only	0.59	0.58	0.59	0.54
	text only	0.62	0.59	0.62	0.61
	cognitive only	0.57	0.58	0.55	0.55
	combo - all	0.57	0.56	0.61	0.56
	combo - significant	**0.75**	**0.73**	**0.69**	**0.75**

their own words what problems affect their daily functioning, the description of each problem and the description of how each problem affects the daily functioning of the participant were automatically transcribed via Amazon Transcribe. Afterwards, the transcriptions were provided to an inference model developed from a neural network with two hidden layers, which was previously trained on 168,260 self-reported problems collected from about 25,000 Parkinson's Disease patients, to identify which of 65 symptoms, such as "cognitive slowing" or "memory", were described by a participant [25]. Additionally, the 65 symptoms were automatically grouped into 14 domains, such as "cognition" or "pain". Finally, we counted the number of times a specific problem domain or symptom was reported by each cohort for both general health and personal well being related problem reports. Figure 4 shows that MCI patients reported three times more problems with cognition than healthy controls. They also reported slightly more gait, psychiatric, and sleep problems. Figure 5 shows that MCI patients reported nearly 2 times more problems with anxiety or worry and speech, nearly 4 times more problems with memory, and 6 times more problems with falls than healthy controls. These results are not surprising because anxiety is common in MCI [26] and the number of falls has been shown to be higher for people with MCI due to impaired balance, gait, and executive functioning [27] (although this is still an ongoing area of research [28]).

6. Discussion

The results of the feasibility analysis show that an elderly population with MCI was not only able to interact with the dialog platform and successfully complete remote assessments, but in fact the majority of participants felt engaged, liked their experience, and would be happy to use such an assessment platform

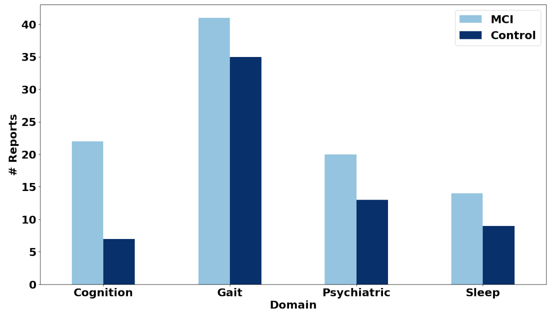

Figure 4: *Overview of the domains of the self-reported problems affecting the daily functioning of participants. Only domains that showed a difference of more than 5 reports between cohorts were included.*

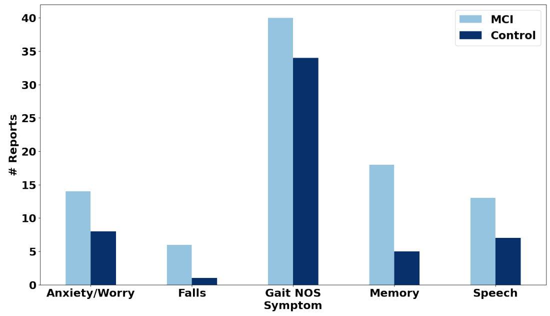

Figure 5: *Overview of the symptoms of the self-reported problems affecting the daily functioning of participants. Only symptoms that showed a difference of more than 5 reports between cohorts were included.*

often or at least sometimes. Additionally, the conducted analytical validation of the automatically extracted biomarkers shows that combining multiple facial, speech, and cognitive features performed best at reliably distinguishing MCI patients from healthy controls with an AUC of 0.75 using a support vector machine classifier, which is similar to results reported in other studies [29, 30, 12]. Finally, the analysis of the symptoms extracted from the self-reported problems shows the utility of asking patients to report their most bothersome problems in their own words.

While the results confirm the feasibility of using remote patient monitoring technology for MCI patients and the utility of combining extracted biomarkers and patient self-reports into a single assessment, there are important caveats and limitations to consider. First, while the number of overall participants was higher than in many other studies, only 10% of the participants were female and the majority of participants was white (representative of the US veteran population), raising the question whether the results are generalizable to more diverse populations. Second, only persons who had an email account and a device with internet access and webcam were invited to participate in the study, introducing a self-selection bias of sorts for veterans who were relatively more tech-savvy than their peers. Finally, while the current work did not have access to measures of cognitive impairment degree, such as MoCA or MMSE scores, future work will investigate the extent to which such multimodal biomarkers can be used to characterize MCI severity, an important step towards demonstrating clinical utility. This includes research into additional feature sets, feature selection methods and machine learning models to improve classification performance and accurately track longitudinal progression of the disease, with applications to clinical trials and clinical care.

7. References

[1] S. Gauthier, B. Reisberg, M. Zaudig, R. C. Petersen, K. Ritchie, K. Broich, S. Belleville, H. Brodaty, D. Bennett, H. Chertkow *et al.*, "Mild cognitive impairment," *The lancet*, vol. 367, no. 9518, pp. 1262–1270, 2006.

[2] K. M. Langa and D. A. Levine, "The diagnosis and management of mild cognitive impairment: A clinical review," *Jama*, vol. 312, no. 23, pp. 2551–2561, 2014.

[3] K. Ritchie and J. Touchon, "Mild cognitive impairment: conceptual basis and current nosological status," *The Lancet*, vol. 355, no. 9199, pp. 225–228, 2000.

[4] H. Posner, R. Curiel, C. Edgar, S. Hendrix, E. Liu, D. A. Loewenstein, G. Morrison, L. Shinobu, K. Wesnes, and P. D. Harvey, "Outcomes assessment in clinical trials of alzheimer's disease and its precursors: readying for short-term and long-term clinical trial needs," *Innovations in clinical neuroscience*, vol. 14, no. 1-2, p. 22, 2017.

[5] R. C. Petersen and S. Negash, "Mild cognitive impairment: An overview," *CNS spectrums*, vol. 13, no. 1, pp. 45–53, 2008.

[6] L. Boise, M. B. Neal, and J. Kaye, "Dementia assessment in primary care: results from a study in three managed care systems," *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, vol. 59, no. 6, pp. M621–M626, 2004.

[7] U. Petti, S. Baker, and A. Korhonen, "A systematic literature review of automatic alzheimer's disease detection from speech and language," *Journal of the American Medical Informatics Association*, vol. 27, no. 11, pp. 1784–1797, 2020.

[8] V. Boschi, E. Catricala, M. Consonni, C. Chesi, A. Moro, and S. F. Cappa, "Connected speech in neurodegenerative language disorders: a review," *Frontiers in psychology*, vol. 8, p. 269, 2017.

[9] S. Luz, F. Haider, S. de la Fuente Garcia, D. Fromm, and B. MacWhinney, "Alzheimer's dementia recognition through spontaneous speech," *Frontiers in computer science*, vol. 3, p. 780169, 2021.

[10] I. Martínez-Nicolás, T. E. Llorente, F. Martínez-Sánchez, and J. J. G. Meilán, "Ten years of research on automatic voice and speech analysis of people with alzheimer's disease and mild cognitive impairment: a systematic review article," *Frontiers in Psychology*, vol. 12, p. 620251, 2021.

[11] V. Ramanarayanan, A. C. Lammert, H. P. Rowe, T. F. Quatieri, and J. R. Green, "Speech as a biomarker: Opportunities, interpretability, and challenges," *Perspectives of the ASHA Special Interest Groups*, vol. 7, no. 1, pp. 276–283, 2022.

[12] V. Vincze, G. Gosztolya, L. Tóth, I. Hoffmann, G. Szatlóczki, Z. Bánréti, M. Pákáski, and J. Kálmán, "Detecting mild cognitive impairment by exploiting linguistic information from transcripts," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2016, pp. 181–187.

[13] M. Asgari, J. Kaye, and H. Dodge, "Predicting mild cognitive impairment from spontaneous spoken utterances," *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, vol. 3, no. 2, pp. 219–228, 2017.

[14] K. C. Fraser, K. Lundholm Fors, M. Eckerström, F. Öhman, and D. Kokkinakis, "Predicting mci status from multimodal language data using cascaded classifiers," *Frontiers in aging neuroscience*, vol. 11, p. 205, 2019.

[15] V. Berisha, C. Krantsevich, G. Stegmann, S. Hahn, and J. Liss, "Are reported accuracies in the clinical speech machine learning literature overoptimistic?" in *Proc. Interspeech 2022*, 2022, pp. 2453–2457.

[16] C. Flint, M. Cearns, N. Opel, R. Redlich, D. M. Mehler, D. Emden, N. R. Winter, R. Leenings, S. B. Eickhoff, T. Kircher *et al.*, "Systematic misestimation of machine learning performance in neuroimaging studies of depression," *Neuropsychopharmacology*, vol. 46, no. 8, pp. 1510–1517, 2021.

[17] I. Shoulson, L. Arbatti, A. Hosamath, S. W. Eberly, and D. Oakes, "Longitudinal cohort study of verbatim-reported postural instability symptoms as outcomes for online parkinson's disease trials," *Journal of Parkinson's Disease*, vol. 12, no. 6, pp. 1969–1978, 2022.

[18] V. Ramanarayanan, D. Pautler, L. Arbatti, A. Hosamath, M. Neumann, H. Kothare, O. Roesler, J. Liscombe, A. Cornish, D. Habberstad, V. Richter, D. Fox, D. Suendermann-Oeft, and I. Shoulson, "When Words Speak Just as Loudly as Actions: Virtual Agent Based Remote Health Assessment Integrating What Patients Say with What They Do," in *Proc. INTERSPEECH 2023*, 2023, pp. 678–679.

[19] V. Ramanarayanan, "Multimodal technologies for remote assessment of neurological and mental health," *Journal of Speech, Language, and Hearing Research*, pp. 1–8, 2024.

[20] P. Boersma, "Praat, a system for doing phonetics by computer," *Glot. Int.*, vol. 5, no. 9, pp. 341–345, 2001.

[21] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz *et al.*, "The kaldi speech recognition toolkit," in *IEEE 2011 workshop on automatic speech recognition and understanding*, no. CONF. IEEE Signal Processing Society, 2011.

[22] V. Bazarevsky, Y. Kartynnik, A. Vakunov, K. Raveendran, and M. Grundmann, "Blazeface: Sub-millisecond neural face detection on mobile gpus," *arXiv preprint arXiv:1907.05047*, 2019.

[23] Y. Kartynnik, A. Ablavatski, I. Grishchenko, and M. Grundmann, "Real-time facial surface geometry from monocular video on mobile gpus," *arXiv preprint arXiv:1907.06724*, 2019.

[24] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

[25] C. Marras, L. Arbatti, A. Hosamath, A. Amara, K. E. Anderson, L. M. Chahine, S. Eberly, D. Kinel, S. Mantri, S. Mathur *et al.*, "What patients say: large-scale analyses of replies to the parkinson's disease patient report of problems (pd-prop)," *Journal of Parkinson's Disease*, no. Preprint, pp. 1–11, 2023.

[26] L. Ma, "Depression, anxiety, and apathy in mild cognitive impairment: current perspectives," *Frontiers in aging neuroscience*, vol. 12, p. 9, 2020.

[27] T. Liu-Ambrose, M. C. Ashe, P. Graf, B. L. Beattie, and K. M. Khan, "Mild cognitive impairment increases falls risk in older community-dwelling women," *Physical therapy*, vol. 88, no. 12, p. 1482, 2008.

[28] V. Leroy, V. Martinet, O. Nunkessore, C. Dentel, H. Durand, D. Mockler, F. Puisieux, B. Fougère, and Y. Chen, "The nebulous association between cognitive impairment and falls in older adults: a systematic review of the literature," *International journal of environmental research and public health*, vol. 20, no. 3, p. 2628, 2023.

[29] R. Nagumo, Y. Zhang, Y. Ogawa, M. Hosokawa, K. Abe, T. Ukeda, S. Sumi, S. Kurita, S. Nakakubo, S. Lee *et al.*, "Automatic detection of cognitive impairments through acoustic analysis of speech," *Current Alzheimer Research*, vol. 17, no. 1, pp. 60–68, 2020.

[30] V. Sanborn, R. Ostrand, J. Ciesla, and J. Gunstad, "Automated assessment of speech production and prediction of mci in older adults," *Applied Neuropsychology: Adult*, vol. 29, no. 5, pp. 1250–1257, 2022.