Predicting Mini-Mental Status Examination Scores through Paralinguistic Acoustic Features of Spontaneous Speech

Ziyang Fu[1], Fasih Haider[1], Saturnino Luz[1]

***Abstract*— Speech analysis could provide an indicator of cognitive health and help develop clinical tools for automatically detecting and monitoring cognitive health progression. The Mini Mental Status Examination (MMSE) is the most widely used screening tool for cognitive health. But the manual operation of MMSE restricts its screening within primary care facilities. An automatic screening tool has the potential to remedy this situation. This study aims to assess the association between acoustic features of spontaneous speech and assess whether acoustic features can be used to automatically predict MMSE score. We assessed the effectiveness of paralinguistic feature set for MMSE score prediction on a balanced sample of DementiaBank's Pitt spontaneous speech dataset, with patients matched by gender and age. Linear regression analysis shows that fusion of acoustic features, age, sex and years of education provides better results (mean absolute error, MAE = 4.97, and $R^2 = 0.261$) than acoustic features alone (MAE = 5.66 and $R^2 = 0.125$) and age, gender and education level alone (MAE of 5.36 and $R^2 = 0.17$). This suggests that the acoustic features of spontaneous speech are an important part of an automatic screening tool for cognitive impairment detection.**

***Clinical relevance*— We hereby present a method for automatic screening of cognitive health. It is based on acoustic information of speech, a ubiquitous source of data, therefore being cost-efficient, non-invasive and with little infrastructure required.**

I. INTRODUCTION

Cognitive ability, also known as cognitive functioning or cognitive intelligence, refers to brain-based capabilities of processing and applying information. To be more specific, these capabilities involve learning, abstract thinking, reasoning, remembering, problem solving, decision making, attention, comprehending complex ideas, and so on ([1], [2], [3]. Cognitive impairment is a great threat to public health. According to Sachdev et Al., the overall prevalence of Mild Cognitive Impairment (MCI) ranges from 5.9% to 12% based on different diagnostic criteria [4]. Regarding the aspect of dementia, it is the fifth largest cause of death, with 2.4 million deaths (4.4% of total death) in 2016. Furthermore, the dementia population is expanding rapidly. In 2016, 43.8 million people were living with dementia, with the number doubling during the past 25 years.

The Mini Mental Status Exam (MMSE) is an examination for evaluating the cognitive status of patients, which consists of multiple test items originally grouped into "*orientation, memory, attention, naming, follow verbal and written commands, write a sentence spontaneously, and copy a complex polygon*" [5]. The performance in each test item is scored and these scores are added together. The total achievable score of MMSE is 30 points. Patients with a MMSE score of under 23 are conventionally considered to have cognitive impairment [6]. However, some studies have suggested a higher threshold (e.g. 26 point) may be better for sensitivity and specificity of identification of cognitive impairment [7], [8].

Changes in speech and language among people with cognitive impairment have been noticeable in patients with MCI and dementia [9], [10]. Appell et al. reported word finding to be difficult, represented by a high incidence of circumlocutions [11]. In addition, Ahmed et al. found that two-thirds of patients with MCI suffered various degrees of decline in connected speech [12]. Several studies provide evidence that speech rate (i.e. phonemes per second) and the amount of pausing during speech had a significant correlation with the severity of cognitive impairment both in narrative and dialogical speech [13], [14], [10], [15]. Furthermore, the findings of Bayles et al. suggested that impairment of language ability may have a linear declining trend and becomes severe with the development of cognitive impairment [16]. As a result, language impairment is considered an important symptom contributing to the diagnosis of cognitive impairment [17], [18].

Picture description is a common way to generate connected speech and is widely used in epidemiological studies to assess language ability. For example, The Cookie Theft picture has been a commonly used picture in many studies [12], [19]. A similar way to picture description, aiming at generating connected speech, is asking patients to recall and describe a past experience, e.g. one of their happiest experiences [20]. This group of tasks mainly focuses on semantic content or syntactic complexity, which have a good ability of capturing the global progression of linguistic impairment through successive screening [12]. Many studies based on picture description have suggested different words' selection between patients with cognitive impairment and healthy controls. It was found that patients with cognitive impairment produce more meaningless content [21], [19], which can account for finding that patients need a larger number of words to describe a picture [22]. As well as the number of words, the speech produced by patients tends to contain a larger proportion of nouns, pronouns, and shorter and more commonly used words [21], [23]. However, language (content) based diagnosis have methodological limitations and do not always generalise well across languages [24], [25].

To our knowledge only a very few studies have tried to predict MMSE scores particularly using only the acoustic

[1] Z. Fu, F. Haider and S. Luz are with the Usher Institute, Edinburgh Medical School, The University of Edinburgh, UK, {s1818665, fasih.haider, s.luz}@ed.ac.uk

information. Only one study used acoustic features in combination with Lexicosyntactic and semantic features (along with manually transcribed content) [26], obtaining a MAE of 3.83 but no results are reported separately for acoustic features only. In addition to linguistic variables, Mendes et al. established a prediction equation with age, sex and education as independent variables. However, the accuracy of this prediction equation has not been reported. The result of this study showed that all three variables are important predictors, because 38% of the total variance of the MMSE scores can be explained by these three variables [27]. We have highlighted the global burden of cognitive impairment and the benefits of screening for cognitive impairments. An automatic screening tool based on speech might provide an economical and scalable screening method. Speech-based methods have shown good performance on Alzheimer dementia classification [15], [10], but only one study used speech-related features to predict the results of the MMSE [26], which is the most widely-used screening tool in primary care facilities. However the limitation of previous study [26] is that it relies on manual transcription of data.

The work presented in this paper contributes to research into MMSE prediction by evaluating and demonstrating the potential of acoustic features based on the ComParE feature set [28] and their fusion with age, gender and level of education, for automatic MMSE prediction. This is, to the best of our knowledge, the first empirical attempt to use acoustics feature sets as "digital biomarkers" for automatic MMSE score prediction.

II. Experimentation

A. Dataset

The Pitt corpus was gathered longitudinally between 1983 and 1988 on a yearly basis as part of the Alzheimer Research Program at the University of Pittsburgh [29]. Participants are categorised into three groups such as dementia, control (i.e. healthy), and unknown. All participants were required to be above 44 years of age, have at least seven years of education, have no history of nervous system disorders or be taking neuroleptic medication, have an initial MMSE score of 10 or more and be able to provide informed consent. Extensive neuropsychological and physical assessments conducted on the participants are also included [30].

The Pitt Corpus contains participants' speech data collected by the Alzheimer and Related Dementia Study at the University of Pittsburgh School of Medicine on the following tasks: a picture description task in which the participant is asked to describe, verbally in their own words, a picture, a word fluency task, a story recall task, and a sentence construction task.

We specifically chose the picture (shown in Figure 1) description task sample for the present study, as it encompasses spontaneously generated narrative speech. The data include MMSE scores for control (cognitively normal) participants, patients diagnosed with MCI, and patients diagnosed with probable Alzheimer's dementia (AD). The selected participants are matched for age and gender (Table I). The resulting

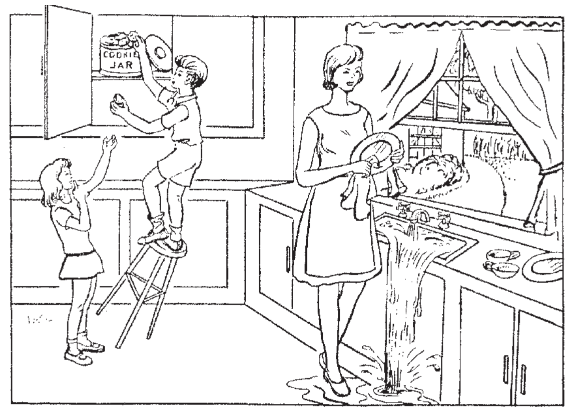

Fig. 1. Cookie Theft picture [31]

dataset was segmented for voice activity using a signal energy threshold. We set the energy threshold parameter to 65dB with a maximum duration of 10 seconds for a speech segment. The segmented dataset contains 2033 speech segments from 82 non-AD subjects and 2043 speech segments from 82 AD subjects. The average number of speech segments produced by each participant in their descriptions was 24.86 (standard deviation $sd = 12.84$). Audio volume was normalized across all speech segments to control for variation caused by recording conditions, such as microphone placement.

TABLE I

Basic characteristics of the patients in each group (AD/non-AD)

Age Interval	AD/MCI Male	AD/MCI Female	Control Male	Control Female	Mean (SD) MMSE
$[50, 55)$	2	1	2	1	27.8 (2.6)
$[55, 60)$	7	8	7	8	23.8 (6.6)
$[60, 65)$	4	9	4	9	23.7 (7.3)
$[65, 70)$	10	14	10	14	24.7 (6.2)
$[70, 75)$	9	11	9	11	22.9 (7.4)
$[75, 80)$	4	3	4	3	23.78 (6.7)
Total	36	46	36	46	
Mean (SD)	19.9 (6.0)	18.3 (6.1)	29.0 (1.0)	29.1 (1.3)	24.0 (6.7)

B. Feature Extraction and Selection

We have extracted ComParE feature set from the speech segments. The *ComParE 2013* [28] feature set includes energy, spectral, Mel-Frequency Cepstral Coefficients (MFCC), and voicing related Low-Level Descriptors (LLDs). LLDs include logarithmic harmonic-to-noise ratio, voice quality features, Viterbi smoothing for F0, spectral harmonicity and psychoacoustic spectral sharpness. Statistical functionals are also computed, bringing the total to 6,373 features. Pearson's correlation test was performed on the whole dataset to remove acoustic features that were significantly correlated with duration (when $R > 0.2$). "Dummy variables" caused by homogeneity of LLD groups were also detected and removed. Detection of dummy variables employed the Variance Inflation Factor (VIF) to detect the multiple collinearity in a linear model. As a result, a linear model was established to detect the dummy features. The equation of this model is shown below:

$$F_i = \beta_0 + \sum_{i=1}^{n} \beta_i F_i + \epsilon \tag{1}$$

where n is the number of features after removing duration related features, F_i is the ith acoustic features after removing duration-related features, β is the estimated regression coefficients, ϵ is the random error. The value of the VIF was also calculated. According to Bollinger, a rule of thumb is that multicollinearity exists when the VIF is larger than 10 [32]. In other words, a feature can be considered as a dummy feature when VIF is larger than 10.

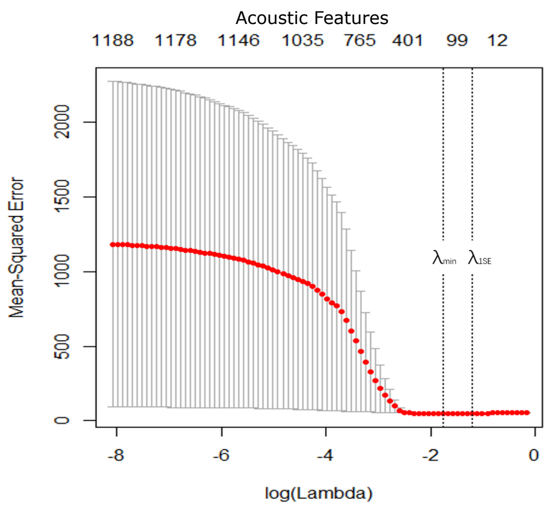

Fig. 2. Mean Squared Error and Number of Variables corresponds to Lambda (λ, the penalisation factor)

Then we applied the least absolute shrinkage and selection operator (LASSO) method to select final acoustic features that might be associated with MMSE scores. We determined the value of non-negative tuning parameter (λ) based on the mean squared error. After λ was determined, the features with a coefficient unequal to 0 were selected as the predictors of MMSE scores. Figure 2 shows the result of the feature selection with LASSO. The vertical axis represents the value of the mean squared error. A lower value of mean squared error indicates a better model. The lower horizontal axis is the value of $log\lambda$, while the upper horizontal axis is the number of features of non-zero coefficient determined by each specific λ. The left dotted vertical line (λ_{min}) represents the value of log λ when the mean squared error has the lowest value, while the right dotted vertical line (λ_{1SE}) represents the value of log λ, which determines the smallest number of features of non-zero coefficient within the range of 1 standard error of lowest mean squared error. Therefore, λ_{min} corresponds to the model that best fits our dataset, but it has a higher risk of over fitting, while λ_{min}, despite the slightly higher mean squared error, has a better resistance to over fitting. As a result, we have 52 features left for further analysis using λ_{1SE}. The selected features contain 5 energy-related features, 43 spectral-related features and 4 voicing-related features. More than half of the selected features (27 of 52) are MFCCs-related features.

III. Results and Discussion

In this study, we repeat 55 times the hypothesis tests for all 52 selected acoustic features and 3 covariates (i.e. age, sex and years of education). As a result, 7 out of 52 acoustic features are considered to have a statistically significant association ($p < 0.001$, adjusted with Bonferroni correction) with MMSE scores. As shown in Table II, these features are all MFCC-related features that belong to the spectral LLD. This suggests MFCCs can be considered as the major group of features that are associated with MMSE scores. MFCCs are a set of acoustic features representing the timbre of a sound, and can be considered as a representation of sound quality. This means that MFCCs' association with MMSE scores might be the association between the sound produced by participants and their MMSE scores. In other words, people with different cognitive statuses might have differences in their vocal timbre. More specifically, the features selected (Table II) describe functionals smoothed by a moving average filter (*sma*), comprising interquartile range between the 50% and the 70% percentiles (*iqr2-3*), percentage of time above 90% of range plus minimum (*upleveltime90*), spectral flatness (*flatness*), root quadratic mean (*rqmean*), and quadratic regression coefficient 3 (*qregc3*).

TABLE II

The Estimated Coefficient (r) of acoustic features with $p < 0.001$

Feature	r	SE	t.value
$mfcc\_sma[1]\_iqr2-3$	-0.07	0.02	-3.46
$mfcc\_sma[1]\_upleveltime90$	-3.13	0.95	-3.31
$mfcc\_sma[7]\_flatness$	4.72	1.16	4.07
$mfcc\_sma[7]\_rqmean$	0.08	0.02	3.82
$mfcc\_sma[10]\_qregc3$	0.03	0.01	4.89
$mfcc\_sma[14]\_rqmean$	-0.15	0.04	-4.00
$mfcc\_sma[14]\_qregc3$	0.05	0.01	4.73

The result of the linear regression model that only includes acoustic features is $R^2 = 0.125$ (Table III), suggesting that the 52 acoustic features that were selected by LASSO account for 12.5% of the variance of MMSE scores. After including sex, age and education, R^2 increased to 0.261. This indicates that inclusion of sex, age and education can improve the prediction of MMSE. However, the value of R^2 is still low, indicating that only about one fourth of the variance of MMSE scores can be explained with the additional consideration of age, sex and years of education. The mean absolute error of the model with only acoustic value was 5.66. After including age, sex and years of education into the model the mean absolute error decreased to 4.97. To further contextualise these results, we classified all speech segments in relation to a clinical criterion for diagnosis of cognitive impairment, setting those with predicted MMSE $<$ 23 as positive instances for cognitive impairment and taking the clinical diagnosis provided by clinicians as ground truth. Table III shows that 70.98% accuracy, 57% sensitivity and 82% specifity can be achieved when all features are used.

TABLE III

LINEAR REGRESSION ANALYSIS RESULTS: ACCURACY(ACC.), SENSITIVITY (SEN.) AND SPECIFICITY (SPE.) IN PERCENTAGE WITH RESPECT TO A CLINICAL CRITERION OF COGNITIVE IMPAIRMENT DETECTION (MMSE$<$ 23) AND CLINICAL GROUND TRUTH USING AGE, GENDER AND EDUCATION LEVEL FOR SPEECH SEGMENTS.

Feature	R^2	MAE	acc.	sen.	spe.
A/G/EL	0.170	5.36	62.82	48.82	78.80
Acoustic	0.125	5.66	65.34	51.65	73.83
Fusion	0.261	4.97	70.98	57.59	82.13

Yancheva et al. also predicted MMSE scores with speech-related features [26] using full dementia bank dataset with longitudinal observations. However we used a sub-corpora of dementia bank dataset without longitudinal observations to avoid bias toward age and gender in our results. Yancheva et al. reported a mean absolute error (MAE) of 3.83 in predicting MMSE. However, they employed lexicosyntactic and semantic features derived from manual transcription, rather than automatically extracted acoustic features as we used in our analysis. In [26], those linguistic features were the main features selected from a group of 477, with acoustic features typically not being among the most relevant. Therefore no quantitative results were reported for acoustic features.

Ambrosini et al. reported an accuracy of 80% while using acoustic (pitch, unvoiced duration, shimmer, pause duration, speech rate), age and educational level features for cognitive decline detection using an Italian dataset of an episodic story telling setting [33]. However, this dataset is less easily comparable to ours, as it is elicited differently, and is not age and gender balanced.

Greater accuracy in MMSE prediction can be obtained if medical imaging data are employed. Zhang et al., for instance, used whole-brain volumes as predictors, and the result of the prediction models with different statistical models showed 7% to 10% average error of prediction [34]. However, whole-brain volumetric data are obtained using magnetic resonance imaging (MRI) devices, which are not typically available in primary care. Therefore, this type of predictors is not helpful to establish an automatic screening tool.

At the other end of the spectrum of data availability, Mendes et al. established a generalized linear model based only on demographic data, namely: age, sex and education [27]. These three demographic variables together accounted for 38% of variance of the participants' MMSE scores. In our model, we also included these three demographic variables, but, even together with acoustic variables, the variance that could be accounted for was still lower than 38%. The different results between the two studies might be because of the difference between the two studies' population. However, both Mendes et al's and our results support that age, sex and years of education are important predictors of MMSE scores. As we have shown, the addition of automatically extracted speech data to demographic data improves MMSE prediction accuracy, and can therefore offer a scalable compromise to complement these easily available patient data for cognitive impairment screening.

A. Limitations

There is limited evidence of an association between the acoustic features and MMSE scores. In order to further interpret the association observed in our study, future research could focus on the relationship between acoustic features and each test item of the MMSE. Secondly, having been collected as part of a research programme that aimed to investigate the natural history of Alzheimer's disease, our dataset contained an inadequate number of young participants. Therefore, future studies could test whether the association between acoustic features and MMSE scores is different among the young population. In our study, our prediction was based on a prediction equation, and its prediction performance required many assumptions, such as linear relationship, independent observation and homoscedasticity [35]. Methods that require fewer assumptions and provide greater generalisation capacity may improve upon the results of our current model [36]. However, no matter which method may be used in future studies, an informative dataset is the precondition of good performance of prediction. Fourthly, our study suggests the possibility of a speech-based automatic screening method for MMSE score prediction. Some researchers have reported the application of a smartphone to collect speech-related features [37]. As a result, future studies could test the feasibility of using a smartphone to automatically screen for cognitive impairments based on the speech-related features collected by the smartphone, and the feasibility of incorporating automatic screening of cognitive impairment into e-health services.

IV. CONCLUSION

Our results suggest that there is a significant association between acoustic features (mainly MFCCs) and MMSE scores. Although this association is not the dominant explanation for the variance of MMSE scores, acoustic features contribute greatly to MMSE score prediction when fused with age, gender and education level features.

In future work, we wish to extend this study further and apply it to spontaneous dialogue data, which we are currently collecting following the PREVENT-ED protocol [38]. PREVENT-ED participants are healthy adults with a comprehensive risk profile (genetics, cognitive assessments and family history of AD), imaging (PET, MRI) and biomarker data, as well as spontaneous speech, are collected longitudinally, thus offering a good opportunity for the investigation of possible relationships between acoustic features of spontaneous speech and clinical markers of cognitive changes.

ACKNOWLEDGMENTS

This research has received funding from the European Union's Horizon 2020 research programme, under grant agreement No 769661, towards the SAAM project. We acknowledge B. MacWhinney (DementiaBank) for hosting and sharing the databases used in this research.

References

[1] G. G. Fisher, M. Chacon, and D. S. Chaffee, "Theories of cognitive aging and work," in *Work Across the Lifespan*. Elsevier, 2019, pp. 17–45.

[2] L. S. Gottfredson, "Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography," 1997.

[3] D. S. Ones, S. Dilchert, and C. Viswesvaran, "Cognitive abilities," *The Oxford handbook of personnel assessment and selection*, p. 179, 2012.

[4] P. S. Sachdev, D. M. Lipnicki, N. A. Kochan, J. D. Crawford, A. Thalamuthu, G. Andrews, C. Brayne, F. E. Matthews, B. C. Stephan, R. B. Lipton *et al.*, "The prevalence of mild cognitive impairment in diverse geographical and ethnocultural regions: the cosmic collaboration," *PloS one*, vol. 10, no. 11, p. e0142388, 2015.

[5] M. F. Folstein, S. E. Folstein, and P. R. McHugh, "“mini-mental state": a practical method for grading the cognitive state of patients for the clinician," *Journal of psychiatric research*, vol. 12, no. 3, pp. 189–198, 1975.

[6] J. R. Cockrell and M. F. Folstein, "Mini-mental state examination," *Principles and practice of geriatric psychiatry*, pp. 140–141, 2002.

[7] W. Kukull, E. Larson, L. Teri, J. Bowen, W. McCormick, and M. Pfanschmidt, "The mini-mental state examination score and the clinical diagnosis of dementia," *Journal of clinical epidemiology*, vol. 47, no. 9, pp. 1061–1067, 1994.

[8] M. Lezak, "Neumpsychologicalassessment (3rd cd.)," 1995.

[9] V. Taler and N. A. Phillips, "Language performance in Alzheimer's disease and mild cognitive impairment: a comparative review," *Journal of clinical and experimental neuropsychology*, vol. 30, no. 5, pp. 501–556, 2008.

[10] S. Luz, "Longitudinal monitoring and detection of Alzheimer's type dementia from spontaneous speech data," in *Computer-Based Medical Systems (CBMS), 2017 IEEE 30th International Symposium on*. IEEE, 2017, pp. 45–46.

[11] J. Appell, A. Kertesz, and M. Fisman, "A study of language functioning in Alzheimer patients," *Brain and language*, vol. 17, no. 1, pp. 73–91, 1982.

[12] S. Ahmed, A.-M. F. Haigh, C. A. de Jager, and P. Garrard, "Connected speech as a marker of disease progression in autopsy-proven Alzheimer's disease," *Brain*, vol. 136, no. 12, pp. 3727–3737, 2013.

[13] I. Hoffmann, D. Nemeth, C. D. Dye, M. Pákáski, T. Irinyi, and J. Kálmán, "Temporal parameters of spontaneous speech in Alzheimer's disease," *International journal of speech-language pathology*, vol. 12, no. 1, pp. 29–34, 2010.

[14] L. Tóth, I. Hoffmann, G. Gosztolya, V. Vincze, G. Szatlóczki, Z. Bánréti, M. Pákáski, and J. Kálmán, "A speech recognition-based solution for the automatic detection of mild cognitive impairment from spontaneous speech," *Current Alzheimer Research*, vol. 15, no. 2, pp. 130–138, 2018.

[15] S. Luz, S. de la Fuente, and P. Albert, "A method for analysis of patient speech in dialogue for dementia detection," in *Resources for processing of linguistic, paralinguistic and extra-linguistic data from people with various forms of cognitive impairment*, D. Kokkinakis, Ed. ELRA, May 2018, pp. 35–42.

[16] K. A. Bayles, C. K. Tomoeda, and M. W. Trosset, "Relation of linguistic communication abilities of Alzheimer's patients to stage of disease," *Brain and language*, vol. 42, no. 4, pp. 454–472, 1992.

[17] G. M. McKhann, D. S. Knopman, H. Chertkow, B. T. Hyman, C. R. Jack Jr, C. H. Kawas, W. E. Klunk, W. J. Koroshetz, J. J. Manly, R. Mayeux *et al.*, "The diagnosis of dementia due to Alzheimer's disease: Recommendations from the national institute on aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease," *Alzheimer's & dementia*, vol. 7, no. 3, pp. 263–269, 2011.

[18] R. A. Sperling, P. S. Aisen, L. A. Beckett, D. A. Bennett, S. Craft, A. M. Fagan, T. Iwatsubo, C. R. Jack Jr, J. Kaye, T. J. Montine *et al.*, "Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the national institute on aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease," *Alzheimer's & dementia*, vol. 7, no. 3, pp. 280–292, 2011.

[19] K. D. Mueller, R. L. Koscik, L. S. Turkstra, S. K. Riedeman, A. LaRue, L. R. Clark, B. Hermann, M. A. Sager, and S. C. Johnson, "Connected language in late middle-aged adults at risk for Alzheimer's disease," *Journal of Alzheimer's Disease*, vol. 54, no. 4, pp. 1539–1550, 2016.

[20] E. Aramaki, S. Shikata, M. Miyabe, and A. Kinoshita, "Vocabulary size in speech may be an early indicator of cognitive impairment," *PloS one*, vol. 11, no. 5, p. e0155195, 2016.

[21] G. Kavé and M. Goral, "Word retrieval in picture descriptions produced by individuals with Alzheimer's disease," *Journal of clinical and experimental neuropsychology*, vol. 38, no. 9, pp. 958–966, 2016.

[22] 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.

[23] L. Meteyard, E. Quain, and K. Patterson, "Ever decreasing circles: Speech production in semantic dementia," *Cortex*, vol. 55, pp. 17–29, 2014.

[24] A. P.-D. López, M. Calero, and E. Navarro-Gonzalez, "Prediction of cognitive impairment in the elderly by analysing their performance in verbal fluency and in sustained attention," *Revista de neurologia*, vol. 56, no. 1, pp. 1–7, 2013.

[25] W. Muangpaisan, S. Intalapaporn, and P. Assantachai, "Digit span and verbal fluency tests in patients with mild cognitive impairment and normal subjects in thai-community," *Journal of the Medical Association of Thailand*, vol. 93, no. 2, p. 224, 2011.

[26] M. Yancheva, K. Fraser, and F. Rudzicz, "Using linguistic features longitudinally to predict clinical scores for Alzheimer's disease and related dementias," in *Proceedings of SLPAT 2015: 6th Workshop on Speech and Language Processing for Assistive Technologies*, 2015, pp. 134–139.

[27] L. P. d. S. Mendes, F. F. Malta, T. d. O. Ennes, G. A. Ribeiro-Samora, R. C. Dias, B. L. C. Rocha, M. Rodrigues, L. F. Borges, and V. F. Parreira, "Prediction equation for the mini-mental state examination: influence of education, age, and sex," *Fisioterapia e Pesquisa*, vol. 26, no. 1, pp. 37–43, 2019.

[28] F. Eyben, F. Weninger, F. Gross, and B. Schuller, "Recent developments in opensmile, the munich open-source multimedia feature extractor," in *Proceedings of the 21st ACM international conference on Multimedia*. ACM, 2013, pp. 835–838.

[29] J. Corey Bloom and A. Fleisher, "The natural history of Alzheimer's disease," *Dementia*, vol. 34, pp. 405–15, 2000.

[30] J. T. Becker, F. Boiler, O. L. Lopez, J. Saxton, and K. L. McGonigle, "The Natural History of Alzheimer's Disease," *Archives of Neurology*, vol. 51, no. 6, p. 585, 1994.

[31] H. Goodglass and E. Kaplan, *Boston diagnostic aphasia examination booklet*. Lea & Febiger, 1983.

[32] G. Bollinger, "Book review: Regression diagnostics: Identifying influential data and sources of collinearity," 1981.

[33] E. Ambrosini, M. Caielli, M. Milis, C. Loizou, D. Azzolino, S. Damanti, L. Bertagnoli, M. Cesari, S. Moccia, M. Cid *et al.*, "Automatic speech analysis to early detect functional cognitive decline in elderly population," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2019, pp. 212–216.

[34] D. Zhang, "A coefficient of determination for generalized linear models," *The American Statistician*, vol. 71, no. 4, pp. 310–316, 2017.

[35] M. A. Poole and P. N. O'Farrell, "The assumptions of the linear regression model," *Transactions of the Institute of British Geographers*, pp. 145–158, 1971.

[36] Z. Obermeyer and E. J. Emanuel, "Predicting the future—big data, machine learning, and clinical medicine," *The New England journal of medicine*, vol. 375, no. 13, p. 1216, 2016.

[37] I. McGraw, R. Prabhavalkar, R. Alvarez, M. G. Arenas, K. Rao, D. Rybach, O. Alsharif, H. Sak, A. Gruenstein, F. Beaufays *et al.*, "Personalized speech recognition on mobile devices," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 5955–5959.

[38] S. de la Fuente, C. Ritchie, and S. Luz, "Protocol for a conversation-based analysis study: Prevent-ED investigates dialogue features that may help predict dementia onset in later life," *BMJ Open*, vol. 9, no. 3, 2019.