Dement Neurocogn Disord. 2024 Jan;23(1):1-10
https://doi.org/10.12779/dnd.2024.23.1.1
pISSN 1738-1495·eISSN 2384-0757

DND Dementia and Neurocognitive Disorder

Original Article

Harnessing the Power of Voice: A Deep Neural Network Model for Alzheimer's Disease Detection

Chan-Young Park,[1] **Minsoo Kim**,[2] **YongSoo Shim**,[3] **Nayoung Ryoo**,[3] **Hyunjoo Choi**,[4] **Ho Tae Jeong**,[1] **Gihyun Yun**,[2] **Hunboc Lee**,[2] **Hyungryul Kim**,[2] **SangYun Kim**,[5] **Young Chul Youn**[1,6]

[1]Department of Neurology, Chung-Ang University College of Medicine, Seoul, Korea
[2]Research and Development, Baikal AI Inc., Seoul, Korea
[3]Department of Neurology, Eunpyeong St. Mary's Hospital, The Catholic University of Korea, Seoul, Korea
[4]Department of Communication Disorders, Korea Nazarene University, Cheonan, Korea
[5]Department of Neurology, Seoul National University College of Medicine and Seoul National University Bundang Hospital, Seongnam, Korea
[6]Department of Medical Informatics, Chung-Ang University College of Medicine, Seoul, Korea

Received: Oct 11, 2023
Revised: Dec 3, 2023
Accepted: Dec 8, 2023
Published online: Jan 22, 2024

Correspondence to
Youn Chul Youn
Department of Neurology, Chung-Ang University College of Medicine, 84 Heukseok-ro, Dongjak-gu, Seoul 06974, Korea.
Email: neudoc@cau.ac.kr

ORCID iDs
Chan-Young Park
https://orcid.org/0000-0002-0851-0609
Hyunjoo Choi
https://orcid.org/0000-0003-4654-3206
Ho Tae Jeong
https://orcid.org/0000-0001-9228-6912
SangYun Kim
https://orcid.org/0000-0002-9101-5704
Young Chul Youn
https://orcid.org/0000-0002-2742-1759

Funding
This research was supported by grants from the Ministry of SMEs and Startups (Project Number: S3079103) and the

ABSTRACT

Background and Purpose: Voice, reflecting cerebral functions, holds potential for analyzing and understanding brain function, especially in the context of cognitive impairment (CI) and Alzheimer's disease (AD). This study used voice data to distinguish between normal cognition and CI or Alzheimer's disease dementia (ADD).
Methods: This study enrolled 3 groups of subjects: 1) 52 subjects with subjective cognitive decline; 2) 110 subjects with mild CI; and 3) 59 subjects with ADD. Voice features were extracted using Mel-frequency cepstral coefficients and Chroma.
Results: A deep neural network (DNN) model showed promising performance, with an accuracy of roughly 81% in 10 trials in predicting ADD, which increased to an average value of about 82.0%±1.6% when evaluated against unseen test dataset.
Conclusions: Although results did not demonstrate the level of accuracy necessary for a definitive clinical tool, they provided a compelling proof-of-concept for the potential use of voice data in cognitive status assessment. DNN algorithms using voice offer a promising approach to early detection of AD. They could improve the accuracy and accessibility of diagnosis, ultimately leading to better outcomes for patients.

Keywords: Voice; Machine Learning; Artificial Intelligence; Alzheimer Disease; Phonetics

INTRODUCTION

Alzheimer's disease (AD), a neurodegenerative disorder, is the leading cause of dementia worldwide.[1] It impacts cognitive abilities, memory, and functional status, progressively eroding an individual's quality of life. The World Health Organization estimated that over 50 million people were living with dementia globally in 2020. As our global population ages, this number is projected to triple by 2050.[2]

Despite the high prevalence and severe impact of AD on individuals and society at large, definitive diagnosis often occurs late in disease progression when therapeutic interventions

Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (Project Number: NRF-2017S1A6A3A01078538).

Conflict of Interest

The authors have no financial conflicts of interest.

Author Contributions

Conceptualization: Youn YC; Data analysis: Youn YC; Funding acquisition: Youn YC; Investigation: Park CY, Jeong HT; Supervision: Kim M, Shim Y, Ryoo N, Choi H, Yun G, Lee H, Kim H, Kim SY; Writing - original draft: Youn YC; Writing - review & editing: Park CY, Youn YC.

are less effective. Traditional diagnostic methods for AD primarily include history taking, neurological examination, cognitive testing, and neuroimaging techniques such as magnetic resonance imaging or positron emission tomography scans.[3,4] However, these methods have significant limitations: 1) they can be expensive and time-consuming; 2) some procedures such as lumbar puncture are invasive; 3) they often require specialized equipment not readily available in all healthcare settings; 4) these approaches might not be practical for early-stage or prodromal AD detection within communities.

In AD diagnostics, various biomarkers, including those derived from plasma, cerebrospinal fluid (CSF), and neuroimaging, have received significant attention. While plasma, CSF biomarkers, and neuroimaging provide direct insights into biological changes occurring in the brain, voice analysis offers a unique window into functional impacts of these changes. Together, these methods can provide a more comprehensive and multidimensional understanding of AD progression. Additionally, voice-based screening, as employed in our study, offers several distinct advantages in the context of AD diagnosis. Especially, voice analysis can be performed remotely, making it a viable option for widespread screening, particularly in underserved or remote areas.

However, voice features captured through advanced computational techniques may reveal subtle changes associated with cognitive decline that are not easily detectable through other biomarkers. These challenges underscore the need for innovative diagnostic approaches that can provide early detection of AD with higher accuracy but lower cost and complexity. Artificial intelligence (AI) offers one such promising avenue. AI algorithms have demonstrated significant potential across various medical fields by transforming raw data into valuable diagnostic insights. Among various biomarkers used for AI-based diagnosis of AD, including imaging data and genetic markers, recent research has suggested that changes in speech patterns might serve as an early sign of cognitive decline linked to AD.[5,6] Speech is a complex task requiring various cognitive processes such as memory retrieval, attention control, and feedback mechanisms, all of which are affected by neurological conditions such as AD.

Alterations in speech patterns include changes in vocabulary usage complexity, semantic density, syntactic simplicity, speech tempo, and pronunciation clarity among others. Such changes can be captured through voice recordings.[7] Harnessing vocal biomarkers offers potential benefits beyond clinical settings alone. Regular voice monitoring within communities could potentially increase awareness about AD risk earlier than before.[8]

Characteristics of voice regulated by cerebral control are hypothesized to differ between individuals with AD and healthy controls. These differences may arise due to neurodegenerative changes that AD imparts on brain structures involved in speech production, such as the cerebral cortex, cerebellum, thalamus, and basal ganglia.[9-11] As these structures are compromised in AD patients,[12] their impairment could manifest as observable changes in voice characteristics. Leveraging machine learning techniques to identify these subtle alterations in voice could enable early screening of AD.

This study aimed to develop a predictive AI algorithm capable of diagnosing AD using vocal biomarkers, an approach less invasive yet potentially more accessible than current diagnostic practices. The goal is to provide a screening tool for communities and clinicians to enable earlier detection, leading to improved patient outcomes through timely intervention.

METHODS

Subjects and dataset

This study was approved by the Institutional Review Board of Chung-Ang University Hospital (registration No. 2022-015-447). The voice dataset contained 221 wav formatted recordings acquired at Chung-Ang University Hospital from February 24th, 2021 to March 18th, 2022. Voice data were obtained from recorded conversations between patients and examiners during a Clinical Dementia Rating (CDR) process. CDR assessments were conducted by 2 trained examiners. Recognizing that the CDR test could vary depending on the administrator and the method used, we took careful measures to standardize the assessment procedure. Audio files were not lossy or compressed, with a Linear 16 sample, a sample rate of 16,000 Hz, and a mono channel. Voices of subjects were manually extracted from the examiner for analysis.

In the dataset, the labeling decision for dementia conformed to probable dementia criteria suggested by the National Institute of Neurological and Communicative Disorders and Stroke and Alzheimer's Disease and Related Disorders Association[13] and the Diagnostic and Statistical Manual of Mental Disorders (DSM)-IV.[14] Subjects were considered to have a mild cognitive impairment (MCI) if they met the following criteria: 1) intact function in activities of daily living, 2) presence of memory complaints, 3) objective cognitive impairment (CI; standard deviation [SD] below education- and age-adjusted norms) in more than one cognitive domain including memory on a comprehensive neuropsychological battery,[15] 4) a CDR of 0.5, and 5) a non-demented case according to the Diagnostic and DSM-IV criteria. Inclusion criteria for normal cognitions were as follows: 1) intact activities of daily living, and 2) no abnormality (within 1.0 SD of education- and age-adjusted norms) on a comprehensive neuropsychological battery.[15]

This study enrolled 3 groups of subjects: 1) 52 subjects having subjective cognitive decline (SCD) with an average age of 72.5 years and an Mini-Mental State Examination (MMSE) score of 25.8; 2) 110 subjects having MCI with an average of 75.9 years and an MMSE score of 24.2; and 3) 59 subjects having Alzheimer's disease dementia (ADD) with an average age of 78.3 years and an MMSE score of 18.2 (**Table 1**). All voice data were anonymized. We obtained voice recordings from each participant through structured conversational tasks involving both the examiner and the participant. During these tasks, the conversation between the examiner and the participant was recorded and subsequently processed for speaker separation. Each conversational recording was split into multiple audio files, where voices of the examiner and the participant were separated. This process allowed us to obtain several distinct voice recordings from each participant. Voice recordings from different participants were not intermixed in training or test dataset. No part of any individual's voice recordings used in the training dataset was included in the testing dataset.

The dataset shown in **Table 2** was divided for predicting 2 main conditions: CI and ADD. The CI encompassed both MCI and ADD. For each condition, there were training and

Table 1. Characteristics of subjects

Variables	No. of subjects	Age	MMSE
NC	52	72.51±9.23	25.75±2.80
MCI	110	75.86±7.05	24.15±3.07
ADD	59	78.34±6.95	18.18±3.29

MMSE: Mini-Mental State Examination, NC: normal cognition, MCI: mild cognitive impairment, ADD: Alzheimer's disease dementia.

https://doi.org/10.12779/dnd.2024.23.1.1

Table 2. Dataset of voice for deep neural network analysis

Variables	Condition	Dataset	No. of data	No. of subjects
Prediction of CI	CI	Training	712	149
		Test	128	20
	NC	Training	360	42
		Test	118	10
Prediction of ADD	ADD	Training	360	48
		Test	73	11
	NC	Training	332	41
		Test	91	11

CI: cognitive impairment (including mild cognitive impairment and Alzheimer's disease dementia), ADD: Alzheimer's disease dementia, NC: normal cognition.

test datasets for both the condition itself and SCD. For CI prediction, the training dataset included 712 data and the test dataset included 128 data. The SCD group for this condition included 360 training data and 118 test data. For ADD prediction, the training dataset included 360 data and the test dataset included 73 data. The SCD group for this condition included 332 training data and 91 test data.

Deep neural networks (DNNs) for diagnostic classification

The study utilized Python programming language with several libraries including scikit-learn, librosa, numpy, scipy, and tensorflow for data processing and model building. The dataset comprised of audio files in .wav format classified into 2 categories: 1) those with CI including MCI and ADD (labelled as 'abnl'); and 2) those with normal cognition (labelled as 'nc').

Firstly, we defined a function named 'extract_features' to extract Mel-frequency cepstral coefficients (MFCCs) and Chroma features from audio files using librosa library (librosa.feature.mfcc() and librosa.feature.chroma_stft()). MFCCs were employed to capture spectral characteristics and provide a compact representation of the audio signal, while chroma features were used to capture tonal characteristics invariant to changes in timbre or articulation.

Data were loaded from 2 directories: one containing files of individuals with CI ('abnl_directory') and another containing files of normal cognition ('nl_directory'). Labels were assigned based on the directory from which each file was loaded.

Feature vectors were extracted for all audio files using the 'extract_features' function. The feature set ('X') and labels ('y') were split into a training set (70%) and a validation set (30%) using the train_test_split method from a scikit-learn library ensuring that the distribution of classes remained similar across both sets. A feed-forward neural network model was built using TensorFlow's Keras API consisting of 4 layers: 1) 3 Dense layers with 'ReLU' activation function followed by 'Dropout' layers for regularization; and 2) and output layer using 'Softmax' activation function suitable for multi-class classification problems. The model was compiled with 'Adam' optimizer with sparse categorical crossentropy loss function since our labels were integers rather than a one-hot encoded, accuracy metric used to evaluate performance during training. Model fitting involved 300 epochs with a batch size of 16 on training data while validating on validation set concurrently.

Performance evaluation included calculating loss & accuracy metrics on validation data after model fitting. These metrics provided an indication about how well our model generalized beyond training data. Predicted class labels were derived by choosing class having the highest probability in the 'Softmax' output for each instance in the validation set. We also evaluated

Table 3. Prediction of CI and ADD

Variables	CI	ADD
Training dataset		
Accuracy	86.21±1.32	80.87±4.90
Test dataset		
Accuracy	69.92±1.68	82.012±1.58
Sensitivity	72.89±5.29	73.01±8.02
Specificity	66.70±5.58	89.23±5.45
AUC	0.71±0.02	0.78±0.04

Values are presented as mean ± standard deviation.
CI: cognitive impairment, ADD: Alzheimer's disease dementia, AUC: area under the curve.

our model against a separate test dataset following similar preprocessing steps as described above but without any involvement in the training process to ensure an unbiased assessment of the final trained model's performance in terms of confusion matrix along with sensitivity & specificity calculations. In addition to its application in predicting CI, the same algorithm was employed to distinguish between individuals with normal cognition and those diagnosed with ADD.

RESULTS

Our AI algorithm aimed to predict CI and ADD using voice data. Performance metrics were computed for 2 distinct binary classification tasks: 1) differentiating between CI and normal cognition, and 2) distinguishing between ADD and normal cognition. These mean values were obtained through 10 trials to ensure robustness of results (**Table 3**).

The DNN showed a high level of accuracy in both training and test datasets for both conditions. It showed a slightly higher accuracy for ADD prediction than for CI prediction in the test dataset. The sensitivity, or true positive rate, was also comparable between the 2 conditions, indicating that the model was able to correctly identify a high proportion of true cases for both CI and ADD. The specificity, or true negative rate, was higher for ADD prediction, suggesting that the model was particularly effective in correctly identifying non-ADD cases.

The area under the curve, a comprehensive metric considering both sensitivity and specificity, was slightly higher for ADD prediction (0.78) than for CI prediction (0.71). This indicates that overall, the model performed slightly better for predicting ADD than for predicting CI. **Fig. 1** shows receiver operating characteristics of 5 trials of the 2 binary classifications.

DISCUSSION

Results of this study demonstrate the potential of using voice data and DNN analysis for predicting CI and ADD. Although our results did not demonstrate the level of accuracy necessary for a definitive clinical tool, they nevertheless provided a compelling proof-of-concept for the potential use of voice data in cognitive status assessment. The DNN model, in particular, showed promising performance when distinguishing between SCD and CI or ADD.

One key advantage of using voice data is its non-invasive nature. If further refined and validated, this approach could enable regular monitoring of individuals at risk without

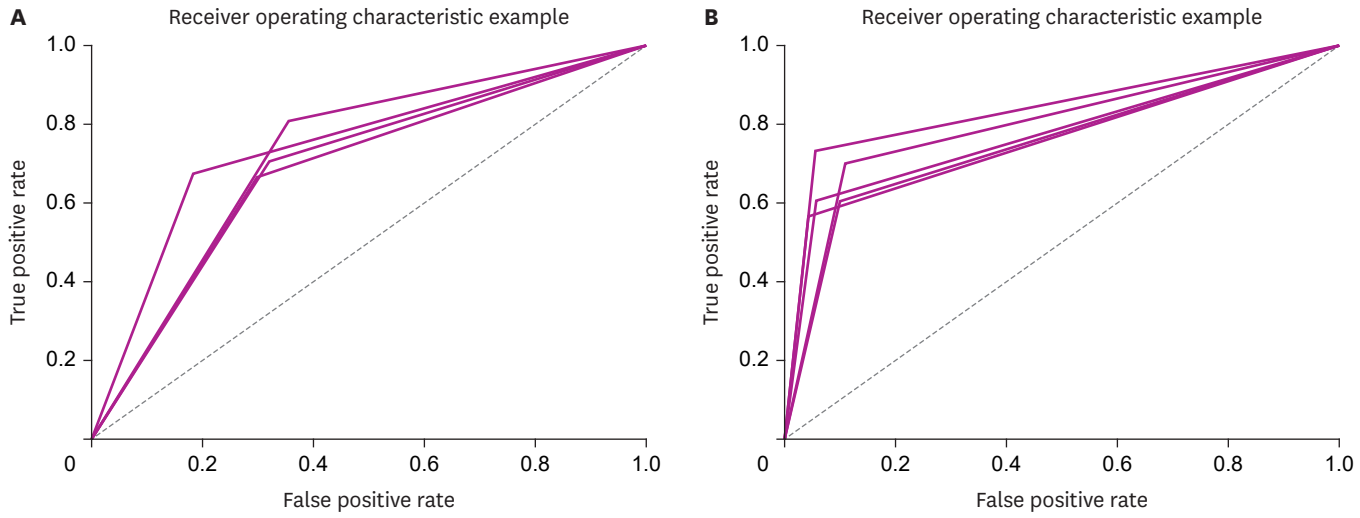

Fig. 1. Receiver operating characteristic curve of deep neural network algorithm using voice data in predicting subjects with cognitive impairment (A) or Alzheimer's disease dementia (B).

requiring physical visits to healthcare facilities, a development that would be especially beneficial for those with limited access to healthcare resources.

Nevertheless, several challenges need to be addressed before such an application can become viable. First, incorporating the A/T/(N) classification could have provided additional insights, particularly for strengthening biomarker-based characterization of study groups. This aspect indeed represents a limitation in our research methodology. In future studies, we aim to integrate such biomarker-based classifications to enhance the comprehensiveness and applicability of our findings in the context of AD diagnostics. In addition, the accuracy of our model needs significant improvement before it can reliably distinguish between different cognitive statuses. This might involve incorporating more complex features into the model or fine-tuning the architecture of our DNN.

The training phase involved optimizing model parameters using various machine learning algorithms such as support vector machines, random forests, and DNNs. Each of these algorithms was evaluated for its performance in differentiating between SCD and CI or ADD (**Appendix 1**). Among these tested algorithms, the DNN demonstrated a superior performance in terms of accuracy, sensitivity, and specificity metrics across both tasks. This indicates that complex pattern recognition capabilities inherent to deep learning models are particularly effective in identifying subtle differences in voice data associated with cognitive status. Consequently, our final model was built using a DNN architecture optimized for this specific task.

Similar to our study, the Framingham Heart Study was previously conducted for detecting dementia using voice recordings with deep learning.[16] However, our methodologies and classification approaches showed notable differences. In the Framingham Heart Study, there was a lack of separation between voices of participant and the examiner in recordings. That study acknowledged that each recording involved 2 speakers. However, that study did not isolate speaker-specific audio. In addition, models of that study processed the entire audio recording at once without performing detailed feature engineering to separate the speaker's voices. Based on our experience with our model, it appears that not implementing speaker

separation could pose significant challenges for AI training and result in a low predictive accuracy. This is a critical observation as it highlights the importance of speaker separation in voice analysis for dementia detection. This suggests that while speaker separation could be a valuable feature in future research, it was not a focus in the Framingham study. In contrast, our research concentrated on extracting and analyzing features from voice recordings using a deep learning model. The Framingham study, meanwhile, employed Long Short-Term Memory for time-series data analysis. It also used image-based (spectrogram) analysis with convolutional neural network, providing a distinct approach to data processing and analysis. Regarding classification, the Framingham study divided participants into 2 categories: normal cognition vs. dementia and non-demented vs. dementia. Our study, however, aimed to differentiate not just between normal cognition and dementia, but also included early stages of CI such as MCI. This distinction in classification targets enabled us to focus more on the spectrum of all CI, including early stages represented by MCI.

The control of voice characteristics such as pitch, formant frequency, and changes in fricative sounds relies on a combination of sensory feedback and feedforward mechanisms.[17,18] This is particularly important considering the complexity of speech production and the need for real-time adjustments to maintain clarity and intelligibility. Sensory feedback involves monitoring the output of speech and making necessary adjustments based on what is heard. For example, if a speaker notices that his/her pitch is too high or low, he/she can adjust it in real time using sensory feedback. Similarly, changes in formant frequency—which affect the timbre or the quality of voice—can also be regulated through this process.[19] On the other hand, feedforward control can predict outcomes of motor commands before sensory feedback is available.[20,21] This predictive mechanism allows speakers to anticipate potential errors in speech production and make proactive corrections.

It is important to note that these processes involve various brain structures including cerebral cortex known to play a critical role in language comprehension and production. For example, the cerebellum is known for its role in motor coordination. The thalamus is involved in relaying sensory signals and basal ganglia are crucial for voluntary movements including speech articulation.[9,10]

Understanding these mechanisms not only provides insights into normal speech production, but also helps us comprehend how neurological conditions like AD might disrupt these processes and lead to observed alterations in voice characteristics.

However, traditional analysis methods for voice data are inherently limited by factors such as the need for standardization and the potential influence of confounding factors. The lack of standardization in voice data due to natural variations in speech across different individuals and influence of variables such as age, gender, and linguistic background is a significant issue.[22] Furthermore, confounding factors such as background noise and recording quality can complicate the analysis process. However, these challenges may be mitigated by using DNN algorithms. Deep learning models are known for their ability to handle high-dimensional data and learn complex patterns from large datasets.[23] By training on diverse examples under various conditions (e.g., different speakers or levels of noise), these models can learn to identify relevant features while disregarding irrelevant variations. Moreover, DNNs excel at automatically extracting useful features from raw data without explicit manual feature engineering, a process known as automatic feature learning or representation learning. This ability can be especially beneficial when dealing with complex and high-dimensional data such as voice recordings.

Therefore, by employing DNN algorithms in our study, we aimed to overcome some limitations associated with traditional analysis methods. Our goal was to develop a more robust model capable of effectively distinguishing between SCD and CI or ADD using voice data despite potential confounding factors. Future research should continue exploring this approach with an aim to improve our diagnostic models using advanced machine learning techniques such as DNNs.

In this study, we extracted voice features using MFCCs and Chroma, both of which are widely used in speech and audio processing. MFCCs provide a compact representation of the spectral shape of an audio signal, making them particularly useful for tasks such as speech recognition and speaker identification.[24] Derived from the Fourier transform of an audio signal mapped to the Mel scale, a perceptual scale that closely approximates human auditory system's response, MFCCs can effectively capture important characteristics of a speech. The computation process involves several steps: calculating the power spectrum of an audio signal, warping this spectrum onto the Mel scale using triangular overlapping windows or filters, taking logarithms, and finally applying discrete cosine transform. This results in a set of coefficients that form a robust representation of the original audio signal. As changes in cerebral control due to AD may affect spectral characteristics captured by MFCCs,[9] these features can potentially contribute to early detection.

Chroma features are another crucial class of audio features extensively used in music information retrieval and speech processing. They can capture tonal characteristics providing a representation invariant to changes in timbre and articulation.[25] Originating from the music theory where 'chroma' refers to twelve different pitch classes within an octave, chroma features can map each frequency bin in an audio spectrogram onto one of twelve chroma bins corresponding to these pitch classes. This results in a compact representation capturing harmonic and melodic characteristics. Similar to the computation process for MFCCs, Chroma involves several steps: computing a spectrogram, mapping each frequency bin into one out twelve chroma bins representing pitch classes, and finally averaging or summing these bins over time to produce a single chroma vector for each frame.[26] In our study, given that changes due AD may affect tonal aspects captured by Chroma,[9] including Chroma as part our feature set could potentially improve our model's performance. Given limitations in understanding the internal mechanisms of our algorithm, it was challenging to precisely identify which voice features, including MFCC or chroma, were frequently selected or given more weight in discriminating CIs from controls. Although these features are known to be valuable in voice analysis, the exact contribution of each feature and the mechanisms through which they enable the differentiation of dementia in our model remain unclear due to the opaque nature of deep learning algorithms. Future research and advancements in model interpretability might provide deeper insights into how these features contribute to the detection of CIs through voice analysis.

In conclusion, while our study did not achieve sufficient accuracy to propose an immediate clinical implementation, it provided an intriguing demonstration of how AI techniques applied to voice data might improve detection and monitoring capabilities in the field of cognitive health.

AVAILABILITY OF DATA AND MATERIAL

All data supporting this study are available from the corresponding author upon reasonable request.

https://doi.org/10.12779/dnd.2024.23.1.1

REFERENCES

1. Selkoe DJ, Lansbury PJ. Alzheimer's disease is the most common neurodegenerative disorder. In: Siegel GJ, Agranoff BW, Albers RW, Fisher SK, Uhler MD, editors. Basic Neurochemistry: Molecular, Cellular and Medical Aspects. 6th ed. Philadelphia: Lippincott-Raven, 1999.
2. World Health Organization. Dementia. Key Facts. Vol 2023. Geneva: World Health Organization, 2023.
3. Knopman DS, DeKosky ST, Cummings JL, Chui H, Corey-Bloom J, Relkin N, et al. Practice parameter: diagnosis of dementia (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology 2001;56:1143-1153. PUBMED | CROSSREF
4. Waldemar G, Dubois B, Emre M, Georges J, McKeith IG, Rossor M, et al. Recommendations for the diagnosis and management of Alzheimer's disease and other disorders associated with dementia: EFNS guideline. Eur J Neurol 2007;14:e1-e26. PUBMED | CROSSREF
5. 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;1:112-124. PUBMED | CROSSREF
6. Themistocleous C, Eckerström M, Kokkinakis D. Voice quality and speech fluency distinguish individuals with mild cognitive impairment from healthy controls. PLoS One 2020;15:e0236009. PUBMED | CROSSREF
7. Garrard P, Maloney LM, Hodges JR, Patterson K. The effects of very early Alzheimer's disease on the characteristics of writing by a renowned author. Brain 2005;128:250-260. PUBMED | CROSSREF
8. Mahon E, Lachman ME. Voice biomarkers as indicators of cognitive changes in middle and later adulthood. Neurobiol Aging 2022;119:22-35. PUBMED | CROSSREF
9. Houde JF, Jordan MI. Sensorimotor adaptation of speech I: compensation and adaptation. J Speech Lang Hear Res 2002;45:295-310. PUBMED | CROSSREF
10. Purcell DW, Munhall KG. Compensation following real-time manipulation of formants in isolated vowels. J Acoust Soc Am 2006;119:2288-2297. PUBMED | CROSSREF
11. Houde JF, Nagarajan SS, Sekihara K, Merzenich MM. Modulation of the auditory cortex during speech: an MEG study. J Cogn Neurosci 2002;14:1125-1138. PUBMED | CROSSREF
12. Braak H, Braak E. Neuropathological stageing of Alzheimer-related changes. Acta Neuropathol 1991;82:239-259. PUBMED | CROSSREF
13. Dubois B, Feldman HH, Jacova C, Dekosky ST, Barberger-Gateau P, Cummings J, et al. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. Lancet Neurol 2007;6:734-746. PUBMED | CROSSREF
14. First MB, Pincus HA. The DSM-IV Text Revision: rationale and potential impact on clinical practice. Psychiatr Serv 2002;53:288-292. PUBMED | CROSSREF
15. Jahng S, Na DL, Kang Y. Constructing a composite score for the Seoul Neuropsychological Screening Battery-Core. Dement Neurocogn Disord 2015;14:137-142. CROSSREF
16. Xue C, Karjadi C, Paschalidis IC, Au R, Kolachalama VB. Detection of dementia on voice recordings using deep learning: a Framingham Heart Study. Alzheimers Res Ther 2021;13:146. PUBMED | CROSSREF
17. ELman JL; JL EL. Effects of frequency-shifted feedback on the pitch of vocal productions. J Acoust Soc Am 1981;70:45-50. PUBMED | CROSSREF
18. Jones JA, Munhall KG. Perceptual calibration of F0 production: evidence from feedback perturbation. J Acoust Soc Am 2000;108:1246-1251. PUBMED | CROSSREF
19. Houde JF, Jordan MI. Sensorimotor adaptation in speech production. Science 1998;279:1213-1216. PUBMED | CROSSREF
20. Pisotta I, Molinari M. Cerebellar contribution to feedforward control of locomotion. Front Hum Neurosci 2014;8:475. PUBMED | CROSSREF
21. Houde JF, Nagarajan SS. Speech production as state feedback control. Front Hum Neurosci 2011;5:82. PUBMED | CROSSREF
22. Konig A, Satt A, Sorin A, Hoory R, Derreumaux A, David R, et al. Use of speech analyses within a mobile application for the assessment of cognitive impairment in elderly people. Curr Alzheimer Res 2018;15:120-129. PUBMED | CROSSREF
23. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436-444. PUBMED | CROSSREF
24. Davis S, Mermelstein P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Trans Acoust Speech Signal Process 1980;28:357-366. CROSSREF
25. Tzanetakis G, Cook P. Musical genre classification of audio signals. IEEE Trans Speech Audio Process 2002;10:293-302. CROSSREF

26. Müller M, Ewert S. Chroma toolbox: Matlab implementations for extracting variants of Chroma-based audio features. In: Proceedings of the International Society for Music Information Retrieval Conference; 2011 October 24–28; Miami. [place unknown]: International Society for Music Information Retrieval, 2011.

Appendix 1. Prediction of CI and ADD in the test dataset using random forest and support vector machine

Variables	Random forest		Support vector machine	
	CI	ADD	CI	ADD
Accuracy	59.36±1.06	65.41±2.31	58.43±1.53	63.13±2.50
Sensitivity	67.81±2.49	60.32±2.97	62.82±4.73	59.22±8.34
Specificity	51.07±0.56	70.68±1.93	54.14±4.26	67.04±5.26

Values are presented as mean ± standard deviation.
CI: cognitive impairment, ADD: Alzheimer's disease dementia.

https://doi.org/10.12779/dnd.2024.23.1.1