JSLHD 2024. Vol. 33, No. 1. 001-011
ISSN 1226-587X / eISSN 2671-7158
Journal of Speech-Language & Hearing Disorders 언어치료연구

ORIGINAL ARTICLE
https://doi.org/10.15724/jslhd.2024.33.1.001
http://jslhd.org/

A Systematic Literature Review on Acoustic of Speech Variables for Measuring Cognitive Function

Mai Ahmed[1*], Soon Bok Kwon[2]

[1] *Dept. of Linguistics, Pusan National University, Doctor's Course, Korean Language and Literature Dept Ain Shams University, Assistant Teacher*
[2] *Dept. of Language and Information, Pusan National University, Professor*

Purpose : Cognitive impairment affects an individual's ability to conduct everyday activities and affects the memory with no cure that can be used to reverse its effects. Therefore, early detection of cognitive impairment is key in dealing with its risks. The purpose of this study is to understand the accuracy of using acoustic variables and machine learning systems in cognitive function evaluation and dementia diagnosis.
Methods : We analyzed a total of 21 papers according to selection criteria to determine research trends. These studies were published between 2013 to April 2023, and they were searched for using three search engines: The National Institutes of Health's National Library of Medicine, Science Direct and BMC Springer.
Results : In terms of acoustics, temporal variables such as pause duration or silent segment duration and ratio were the most reported effective values in cognition function evaluation, followed by pitch and formant characteristics. Also, support vector systems have learning system as the most accurate when used for assessment combined with other models such as LASSO algorithm and Multi-Layer Perceptron (MLP) to reach a classification accuracy of 97% and 93%.
Conclusions : Cognitive function evaluation through acoustic phonetic variable analysis is a promising research field. This points to the need for developing an automated platform system using machine learning for automated evaluation of cognitive function by acoustic variables.

Keywords : Cognitive function measurement, acoustic analysis, dementia, cognitive impairment evaluation, machine learning

Correspondence : Soon Bok Kwon, PhD
E-mail : sbkwon@pusan.ac.kr

Received : November 30, 2023
Revision revised : December 15, 2023
Accepted : January 31, 2024

This work was supported by a journal research promotion for graduate students of Pusan National University (2023).

ORCID
Mai Ahmed
https://orcid.org/0000-0003-2470-9666
Soon Bok Kwon
https://orcid.org/0000-0002-9424-0077

Ⅰ. Introduction

As life expectancy increases, arise a need to study the problems an aging society would face such as the health problems accompanying the old aged. One of these aging health problems is cognitive impairment. Cognitive impairment (CI) isn't only a problem for the patient or family members, it can be a problem for the whole society if not considered properly. That is CI may lead to dementia (DE) which doesn't have a cure and therefore the best intervention for it is at the early stage with memory training. It can be said that early detection is key in maintaining cognitive health.

CI affects the individual's ability to carry out activities of daily living and memory retaining of information. In mild cognitive impairment (MCI) patients' cases, they can present symptoms that are not as significantly detectable as DE patients; meaning they would have more capacity to do their daily tasks and retain more information in their memory (NICE, 2020; Riley et al., 2022). And as it is estimated that half of MCI patients would develop DE later on, early detection of DE through screening of MCI is crucial in controlling DE (Riley et al., 2022). DE in itself is a broad term that covers an array of several diseases such as Alzheimer, which affects about 60% to 70% of DE patients, and other forms such as vascular dementia, dementia with Lewy bodies and frontotemporal dementia (WHO, 2023). People who are more at risk of DE are those who are 65 or older who may suffer of high blood pressure diabetes or obesity, smoke or drink too much alcohol, are physically inactive or socially isolated and suffer from depression (WHO, 2023). According to WHO's "Global action plan on the public health response to dementia 2017~2025," dementia affected 47 million

Copyright 2024 © Korean Speech-Language & Hearing Association.
This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

patients worldwide in 2015, and this number is estimated to rise to 75 million in 2030 and 132 million in 2050. The costs of DE, which are burdened by both the patient and the community or government, were estimated to be at 1.3 trillion USD globally in 2019 (WHO, 2023).

Traditionally brain imaging and face-to-face screening tests were used as tools in diagnosing DE. DE detection tests have also been developed to assess memory capacity, word retrieval and verbal fluency, problem solving and motor abilities. Screening and assessment tests include the Mini-Mental State Examination (MMSE), the General Practitioner assessment of Cognition (GPCOG), Montreal Cognitive Assessment (MoCA), which were developed to assess cognitive function (Riley et al., 2022). These evaluation tools are mainly used in face-to-face clinical settings, but they are also being developed to be used for self or machine evaluation. Development enables remote models to conduct more longitudinal and informational research.

More recent efforts are directed at improvising mobile devices or tablet systems that can be used as a more accessible and cost-effective self-administrated measure, that would also contribute to more informative longitudinal data as it would be used more frequently by the patients than traditional screening examinations (Öhman et al., 2021). Automatic speech analysis has become a research focus of cognitive decline evaluation because it is a non-invasive, non-surgical method of capturing and evaluating subjects' language performance in real time König et al. (2015). Considering that motor activity and linguistic features have been integrated already in existing screening examinations, acoustic or prosodic measurements can be regarded as the novel approach of recent studies. And according to König et al. (2015), human speech signals can be used as biomarkers to detect cognitive decline progression; that is because spontaneous speech in interaction requires the use of multiple memory systems for word retrieval, semantic understanding, and syntax composition (De Looze et al., 2021). This can provide important information that can help evaluate and manage cognitive decline.

Considering acoustic variable analysis use for cognitive function evaluation, previous literature reviews have focused on the applications of machine learning models in cognitive function evaluation, the accuracy and usability of mobile phone system screening, and the comparison between using traditional clinical assessments with novel digitized ones. Among them, Martínez-Nicolás et al. (2021) also aimed at giving a general overview of the methods of which data were collected in neurodegenerative disorders' assessment through voice measures, and how these data are analyzed. Another review by Petti et al. (2020) focused on the relationship between speech features, speech tasks used in collecting data and classification methods used for classifying AD with healthy or MCI with healthy. In this study however, we focus only on the use of acoustic and prosodic features for automating a cognitive capacity classification system. The goal is to analyze the use of acoustic variables considered in previous studies as cognitive deterioration biomarkers, in terms of their accuracy and importance in the evaluation of cognitive decline. In addition, for evaluating the usability of acoustic features in measuring cognitive decline, we investigate the accuracy of the machine learning algorithms that used only acoustic measures in their classifications in the reviewed papers. The reason this paper is focusing on acoustic prosodic variables is that unlike motor skills evaluation, it can be measured easily through a conversation on a phone or tablet. Also, measurement of acoustic features can be done completely automatically with a higher accuracy rate compared to semantic features analysis in some languages where the automation of semantic and syntactic speech analysis is not without faults.

II. Methodology

1. Article search strategy

This study focused on reviewing papers researching the use of phonetic acoustic variables in diagnosing or evaluating cognitive decline. A search was conducted on the National Institutes of Health's National Library of Medicine (PMC) (PubMed), Science Direct and BMC Springer using search terms including: "MCI voice measures", "MCI Alzheimer acoustics", "MCI shimmer", "MCI jitter", "MCI Alzheimer prosodic features", "Alzheimer vocal features", and "MCI acoustic features". The search was focused on articles that allowed open access and have been published in the last 10 years between January 2013 and April 2023. As for the CI types studied in the target papers, papers studying CI types such as AD, frontal lobe dementia, vascular dementia, and Lewy bodies dementia were selected, excluding papers studying dementia caused by Parkinson's disease, primary progressive aphasia, and other causes. In addition, studies that combine acoustic variables with linguistic variables or motor skills in the

cognitive function evaluation model, and studies aimed at proving the accuracy of machine learning methods were also excluded. The criteria for inclusion and exclusion in this review are shown in Table 1 below.

Table 1. Inclusion and exclusion criteria of reviewed article

Criterion	Inclusion	Exclusion
Publishing date	• From Jan. 2013 till April 2023	• Before Jan. 2013
Focus of study (classifications)	• Any combination of classification between: HC, MCI and DE or AD • Articles focusing on prosodic acoustic characteristics as a classification criterion	• Articles focused on Parkinson's disease, primary progressive aphasia or thyroid cancer, diabetes and blood pressure related studies. • Articles combining motor skills and semantic or lexical features with acoustics for classification. • Articles not specifying the acoustic measures used in classification. • Articles focusing only on comparing machine learning systems or proving their accuracies rather than the acoustic variable significance in the study.
Types of articles	• Articles that are peer reviewed • Research articles	• Literature review articles • Conference summaries • Editor comments • Editorials • Case studies • Preliminary study protocols
Language	• Written in English	• Written in any other Language

Note. MCI=mild cognitive impairment; AD=Alzheimer disease; DE=dementia; HC=healthy controls.

2. Selection procedure

After application of filtration for date of publication and access, a total of 4,025 articles were the preliminary search result. After screening articles' titles and abstracts literature reviews, editorials, and conference summaries were excluded; the preliminary screening resulted in a total of 103 articles. A full-text screening of these 103 articles was then done to determine whether the article was focusing on machine learning systems and its efficiency rather than acoustic features, or if the acoustic measures are not specified. The result of the selection procedure was 21 articles which are the focus of this literature review (Figure 1).

Ⅲ. Results

Although this literature review targeted papers published in the last 10 years it has been noticed that about 70% of the literature in this review has been published in the last three years only (2020~2023). This indicates that as a research field the analysis of acoustic variables' use in cognitive assessment is still a relatively new topic with a promising future (Table 2).

1. Acoustic variables

Regarding programs used for feature extraction; PRAAT was the main program used for manual extraction, and in case of automatic extraction other tools such as opensmile toolkit, audacity, automatic speech recognition (ASR) based tool, Voice Activity Detector (VAD), python programs and processing libraries, divas Voice Diagnostics System by XION®. With audacity as the relatively more popular choice between all the tools followed by the opensmile toolkit.

As for analyzed acoustic features, Table 2 shows the variables that were reported to be the most significant in the classification of CI which included silent and pause segments, formant frequency, speech time, jitter, shimmer, MFCC, and GTCC. Figure 3 below shows the variables contributing to the evaluation of cognitive function reported in the literature by category. Variables related to pause and silence interval segments have been reported in 9 papers as significant variables in the screening process, with these variables accounting for the highest percentage at 42.8% of the total of all significant variable reports. Followed by phonation and speaking time-related variables (phoneme rate and articulation rate) and fundamental frequency variables such as the average and *SD* of F0, semitone, central of gravity, asymmetry (skewness), central peak promise and n-PVI accounted for 23.8% each. In four studies, MFCC has been shown to play a significant role in the screening process especially the first 14 MFCCs with emphasis on coefficients 1, 3, 4, 7. Considering the above results, temporal variables can be considered as essential components in the classification set for increasing classification accuracy since most studies have reported that MCI and DE utterances have more silent segments and longer pause times and are characterized by more use of fillers in utterances, all of which are prominent

time-related variables.

2. Machine learning systems

Automatic evaluation relies on machine learning algorithms and statistical models for the classification of data fed to the system. These classifications can be categorical discrimination, or it can go for scaling the data depending on the type of classification modeled in the system. Table 2 shows the machine learning models that were used for cognitive function classification using acoustic variables that were analyzed in the literature. The most noticeable aspect of the data in Table 2 is that the machine learning system that was used most in the literature was SVM. SVM was used in 11 papers and is reported to have a larger accuracy range than other systems, up to 97.7%. Figure 3 describes the frequency of use of machine learning systems in the reviewed literature.

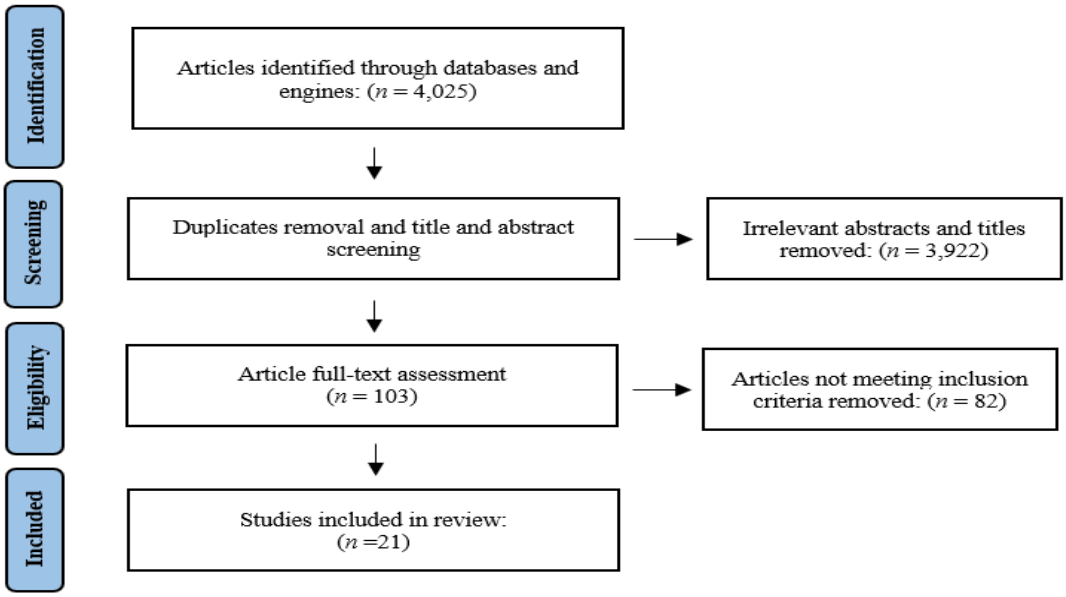
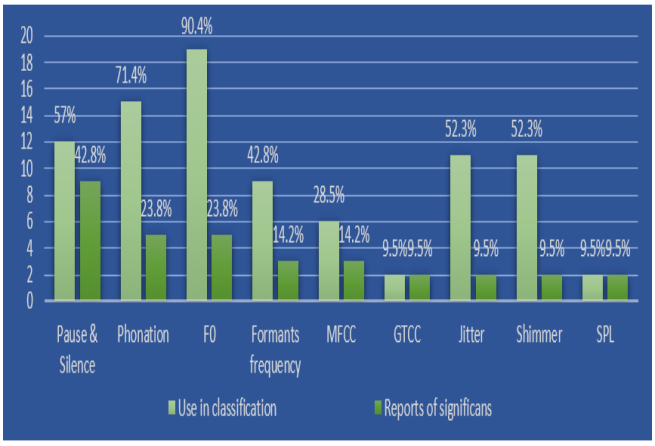
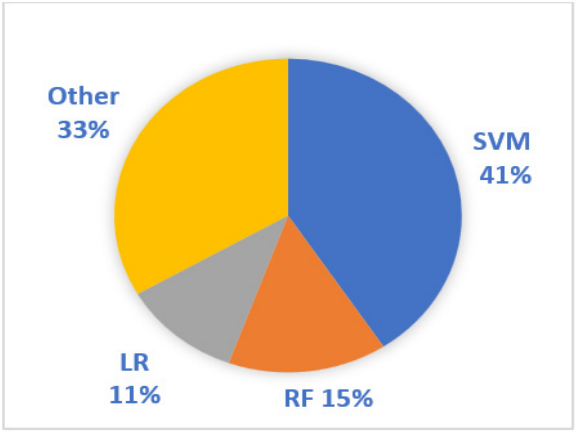


Figure 1. Flow diagram of paper selection process



Note. GTCC=gammatone cepstral coefficients; MFCC=mel-frequency cepstral coefficients; SPL=sound pressure level.

Figure 2. Reports of significance for variables per paper



Note. LR=linear regression; RF=random forest; SVM=support vector machine; Other=reduced error pruning, linear regression, XGBoost, random tree, LightGBM, naive bayes, K-nearst neighbor, linear discriminant analysis, tree Bagger (each used once).

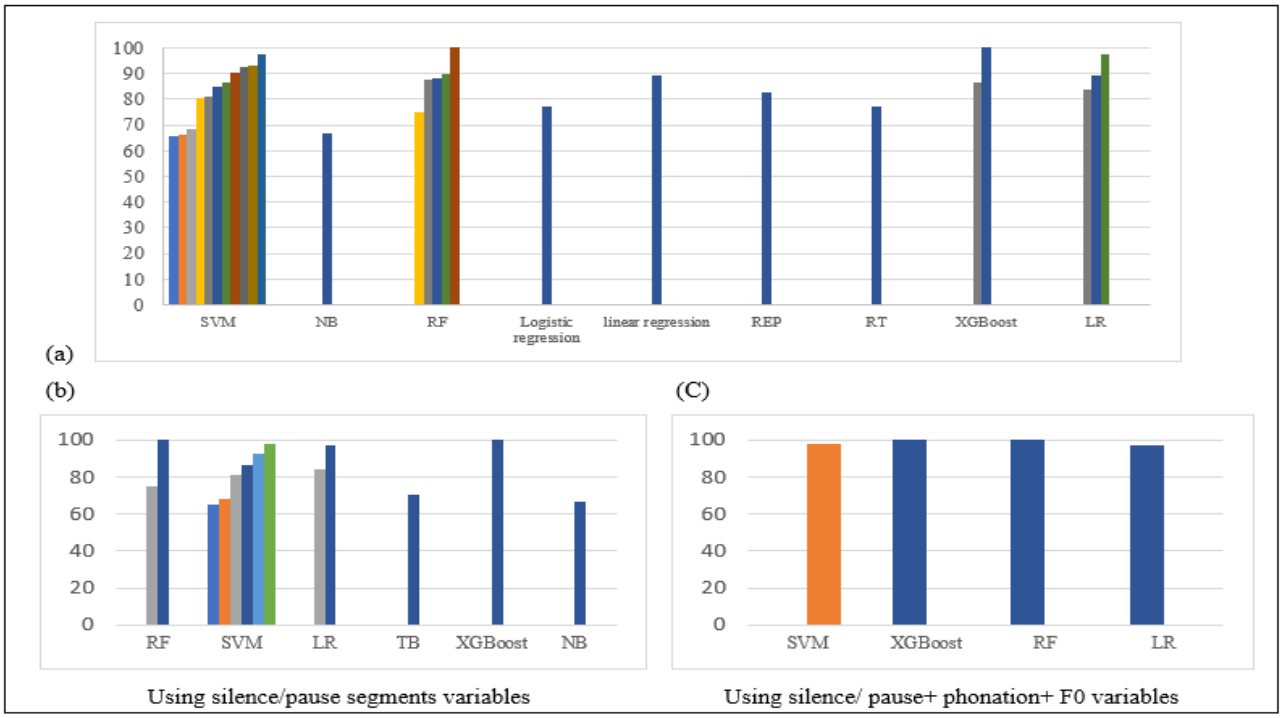
Figure 3. Machine learning systems used in classifications

Table 2. An overview of studies included in the review

Study	Population (Size)	Significant temporal or acoustic measures	MLS	Accuracy of classification
Toth L. et al. (2018)	HC, MCI (84)	1. Total length of silent pauses 2. Length of pauses	1. Naive Bayes 2. RF 3. SVM	Naive Bayes=66.7% RF=75% SVM=65.5%
Nagumo et al. (2020)	HC, MCI, GCI, MCI with GCI (8,779)		Logistic regression of binary classification	AUC: MCI=.61 GCI=.67 MCI with GCI=.77
Gonzalez-Moreira et al. (2015)	HC, Mild DE (20)	1. *SD* of F0 2. Mean of F0	SVM	85%
König et al. (2015)	HC, MCI, AD (64)	Mean of silent segments length	SVM	81%
De Stefano et al. (2021)	HC, MCI, FTD, VD, AD (61)	1. Semitones 2. SPL max	Multiple linear regression analysis	89.3%
Meilán et al. (2020)	nodMCI, preclinical AD (preAD)	1. Skewness (Asymmetry in the spectral feature) 2. nPVI (Variability in articulatory rhythm)		
Kumar et al. (2022)	HC, DE (442subject/recordings)	1. F0 2. log-energy 3. First 4 formants 4. First 14 MFCC 14 GTCC 5. 14 delta-MFCC 14-delta GTCC	SVM RF REP RT	SVM=80.3% RF=87.6% REP=82.6% RT=77.4%
Shimoda et al. (2021)	HC, DE (123)		XGBoost RF LR	Audio file-based prediction: AUCs for XGboost=.863, RF=.882, and LR=.893 Participant-based prediction: AUC for XGboost=1.000, RF=1.000 and LR=.972
Liu et al. (2023)	HC, AD (10, and 393 in training data)	1. Pause feature sequences	LDA DT KNN SVM TB	TB=70.7%
Lin et al. (2020)	Incidental DE and HC (4,849)	1. JitterDDP_sma_de 2. Shorter segments	Cox proportional hazards models with robust sandwich estimators	AUC of 81.2%
Themistocleous et al. (2018)	MCI, HC (55)		SGD optimization algorithm in python	83%
Da Cunha et al. (2022)	LvPPA, AD (16)	Pause rate		
Nishikawa et al. (2022)	HC, MCI	F1, F2 of vowels	SVM (Linear) RF light GBM	SVM (Linear) + ViT_b16[a] =90.4% Random Forest + ViT_b16[a] =89.5% LightGBM + ViT_b16[a] =90.7%
Martínez-Nicolás et al. (2022)	NC, MCI, AD (400)	1. Phonation time 2. Number of pauses in speech		-
López-de-Ipiña et al. (2013)	AD, HC (40)	-	SVM, Multi Layer Perceptron (MLP)	97.7%
Yoshii et al. (2023)	MCI, HC (94)	1. Duration of response time 2. Duration of silent periods 3. Proportion of silent periods	SVM	MMSE related speech=66.0% Everyday conversational speech=68.1%
Sumali et al. (2020)	Depression, DE (120)	1. GTCC coefficients 1, 3, and 12 2. MFCC coefficients 1, 3, 4, 7, and 12	SVM with linear kernel with LASSO algorithm	93.3%±7.7%
Wang et al. (2022)	AD, aMCI, HC (324)	Silence duration	LR	AUC: NC/aMCI=.74 NC/AD=.84 NC/aMCI+AD=.80
Yamada et al. (2021)	MCI, HC (76)	1. MFCC 2. Pause 3. Speech rate 4. Phonation time	SVM	86.40%
Hall et al. (2019)	HC, MCI, AD (44)	1. Jitter and shimmer (Local and APQ3) 2. Silence (Mean and *SD*)	SVM with linear kernel function	HC vs. MCI=82.4 HC vs. DE=92.6%
Themistocleous et al. (2020)	HC, MCI (55)	1. H1-A3 2. Cepstral peak prominence 3. Center of gravity 4. Shimmer 5. Articulation rate 6. Averaged speaking time		

Note. AD=Alzheimer disease; aMCI=amnestic mild cognitive impairment; ANN=artificial neural networks; AUC=area under the curve of the receiver; CN=cognitively normal; DE=dementia; DLB=dementia with Lewy bodies; DT=decision tree; F0=fundamental frequency; FTD=frontotemporal dementia; GCI=global cognitive impairment; GTCC=gammatone cepstral coefficients; HC=healthy control; KNN=K-nearest neighbor; LDA=linear discriminant analysis; LR=logistic regression; LvPPA=logopenic variant of primary progressive aphasia; MCI=mild cognitive impairment; MFCC=mel-frequency cepstral coefficients; MLS=machine learning systems; NC=normal control; nodMCI=nondegenerative MCI; preAD=preclinical AD; REP=reduced error pruning; RF=random forest; RT=random tree; SRP=speech range profile; *SD*=standard deviation; SGD=nesterov stochastic gradient descent; SVM=support vector machine; TB=tree bagger; VD=vascular dementi; GBoost=extreme gradient boosting.

[a] ViT_b16 is a Transformer model that is optimized for image processing.



Note. Diversity in bar colors represent the number of studies using the system and accuracies achieved by each system (data of each model's accuracy for each study is represented in Table 2).
AUC=area under the curve of the receiver; F0=fundamental frequency; LR=logistic regression; NB=naive bayes; RF=random forest; SVM=support vector machine; TB=tree bagger; XGBoost=extreme gradient boostin.

Figure 4. Accuracy and AUC of machine learning in classification

Figure 4 shows the accuracy of cognitive function classification performed by the machine learning systems. Graph (a) shows the accuracy and AUC of all machine learning reported in the literature, while graphs (b) and (c) link the accuracy of machine learning systems to the reportedly significant acoustic variables. Graph (b) presents the accuracy of machine learning systems using pause and silence segments in classification. And graph (c) shows the accuracy of classification using all of pause and silent segments' related variables, phonation and speaking time variables, and fundamental frequency variables. As demonstrated in Figure 4, models that used all most significant variables achieved AUC of more than 90% in the range of accuracy spanning from 65% to 100%.

Ⅳ. Discussion and Conclusion

This study aimed to identify new trends in utilizing non-surgical methods for detecting and classifying cognitive decline through the use of acoustic phonetic variables. The literature review encompassed 21 papers published between 2013 and April 2023, which were selected and analyzed to discern the prevailing patterns in employing machine learning and acoustic measurements for CI assessment; the selected papers were specifically focused on analyzing the use of acoustic variables within the evaluation process. Consequently, studies that integrated semantic, syntactic, or motor ability-related factors with acoustic phonetic variables during the classification process were excluded significant acoustic phonetic variables that contribute to the from the final literature review selection. The primary objective of this review was to identify the most accurate model of machine learning employed and to highlight automatic classification process.

In the context of machine learning models employed for the classification of cognitive decline evaluation, the application of machine learning and the relationships between different models have been elucidated in prior studies. For instance, Agbavor and Liang (2022a) highlighted that healthcare-related applications often make use of Support Vector Classifier (SVC) and logistic regression. Tanaka et al. (2017) conducted a comparison between SVM and logistic regression, concluding that SVM exhibited higher accuracy than the latter. This conclusion finds support in the work of de la Fuente Garcia et al. (2020) in their review. In a comprehensive review encompassing 51 papers on AI approaches to monitoring

AD through speech and language features, SVM, Naive Bayes, Random Forest, and K-Nearest Neighbor emerged as the most frequently used machine learning systems. Notably, SVM consistently outperformed the other models (de la Fuente Garcia et al., 2020).

Agbavor and Liang (2022b) also compared SVC and logistic regression with Random Forest, which they explained has the drawback of not achieving the same level of performance as the previous two models when working with small datasets. While some studies on CI utilize existing databases or corpora (such as dementia bank, AZTIAHORE database, etc.) containing audio files from patients at varying stages of cognitive decline (Javeed et al., 2023), other studies that gather their own data encounter the challenge of limited datasets. Hence, it is reasonable that SVM and logistic regression are favored options. SVM, in particular, emerged as the predominant machine learning model in HC and CI classification. Figure 3 illustrates this dominance, as SVM was employed in 11 out of the 21 studies. This aligns with the findings of another review by Vigo et al. (2022) concerning the classification of AD using language and acoustic features. They noted that SVM was the most widely used classification model in studies connecting speech signals with the disease. They attributed SVM's popularity to its ability to handle non-linear data distributions, which are common in speech-related data (Vigo et al., 2022).

In Figure 3, Random Forest ranks second after SVM as the most frequently used classification model. A similar outcome was reported in Javeed et al.'s (2023) review of research directions in studies utilizing machine learning to predict dementia. They found that SVM was the most frequently utilized machine learning model in the literature, followed by Random Forest. It appears that SVM and Random Forest are the prevailing choices in machine learning systems when employing speech features for the classification of CI. However, in the case of Random Forest, considerations regarding dataset size must be taken into account prior to application, unlike SVM. Furthermore, in the context of binary classifications, both SVM and Random Forest are regarded as optimal tools, with a particular emphasis on SVM's capacity to achieve high accuracy in CI classification (Javeed et al., 2023).

Previous studies and reviews' findings are also consistent with this review's findings. In literature reviewed within this study, SVM emerged as the most commonly utilized model, as mentioned earlier. It was also reported to achieve higher classification accuracies, reaching up to 93% (Sumali et al., 2020) and 97% (López-de-Ipiña et al., 2013). However, it's important to note that the heightened accuracy of the model was achieved by incorporating algorithms such as the LASSO algorithm and neural networks like the Multi-Layer Perceptron (MLP) to enhance the Model's classification criteria. High percentages were not only related to SVM though; in another research conducted by Shimoda et al. (2021), three machine learning models were employed and compared for the classification of dementia using features extracted from everyday conversational speech collected via mobile phones. These models included extreme gradient boosting (XGboost), Random Forest, and Logistic regression. The study reported AUC values of .863, .882, and .893, respectively, for audio-file-based classification, and 1.000, 1.000, and .972, respectively, for participant-based classification. While these outcomes are remarkable, it's worth mentioning that Shimoda et al.'s (2021) study comprised a population of 99 HC and only 24 individuals with DE. The disproportional and limited size of the population might raise concerns about the feasibility of replicating these results with a larger population.

The accuracy of classification depends not only on the machine learning model but also on the acoustic measures utilized in the classifications. Table 3 in the appendix presents a list of acoustic phonetic variables analyzed in each study reviewed in this paper. As shown in Table 3, the list of variables extracted and analyzed for each study differed according to each study. This divergence is also evident in the variation of reports of the most effective variables used in evaluating cognitive function in each study. This variance in the consideration of variables makes cross-study comparisons challenging and points out the need for standardized feature sets (e.g., ComPare, eGeMAPS etc.) to facilitate better comparisons (de la Fuente Garcia et al., 2020). Nonetheless, many studies have attempted to determine the significance of specific acoustic or temporal measures in the classification of cognitive function decline and the relationship between these measures and CI.

As illustrated in Figure 2 the most commonly reported significant acoustic measures in CI classifications are the Temporal related variables. These variables include pause and silence segment's length, total number and ratio of pause and silent segments, in addition to speech duration and phonation related variables. This finding aligns with conclusions drawn in other studies, where it was observed that temporal variables related to pause duration and phonation rate are affected in MCI conditions, and that these temporal variables are considered markers of

cognitive decline even in early stages (Beltrami et al., 2018; Hecker et al., 2022).

Beltrami et al. (2018) explained that the increase in pause duration and decrease in phonation rate result from extended discourse planning, impacted by the deterioration of verbal fluency and lexical retrieval processes as memory is affected by degeneration. Yeung et al. (2021) concluded that word-finding difficulty and incoherence played roles in distinguishing MCI and AD. Word-finding difficulty was inferred from pause duration, phonation duration, word duration, and syntactic complexity, further emphasizing the importance of temporal variables in differentiating not only HC and MCI but also MCI and AD. Boschi et al. (2017) also emphasized the link between discourse planning and temporal features, such as pause and phonation duration, and how these variables indicate CI.

In addition to temporal variables, variables related to fundamental frequency were also found to be significant in the reviewed studies. However, when it comes to fundamental frequency, findings from previous studies differ in their conclusions about its importance in evaluation of CI. For instance, Tanaka et al. (2017) argued that pitch is more related to emotions, depending on emotional state rather than cognitive intelligence. They used this to explain the disparity in their pitch-related results compared to previous studies that associated reduced pitch modulation with AD. It is worth noting here that López-de-Ipiña et al.'s (2013) study, which achieved 97% accuracy in classification between HC and AD, considered emotional temperature in classification. "Emotional Temperature" encompasses several prosodic and paralinguistic features linked to pitch and energy that enable the recognition of emotions in speech. This study integrated the concept of emotional temperature into the classification process, contributing to the higher accuracy achieved. Additionally, machine learning models that combined variables related to pause, silence segments, fundamental frequencies, and phonation/speaking time, as reviewed in this study, demonstrated higher AUC values in classification compared to other models, as illustrated in Figure 4's (c) graph.

Another noteworthy feature in classification is the MFCC and GTCC sets. In studies conducted since 2020, variables like MFCC and GTCC have been utilized in the classification process and have been shown to significantly impact the screening process results. Tang et al. (2022) used MFCC for acoustic classification and argued that the variability of MFCC coefficients' increase and decrease aids in predicting MCI and distinguishing between MCI and HC. This was also mentioned by Voleti et al. (2020), where the variability of MFCC coefficients proved useful in distinguishing between HC and CI over time. MFCC is a large set of features, but studies reviewed in this paper highlighted some specific coefficients and their relation to De. Kumar et al. (2022) highlighted the significance of the first 14 MFCC coefficients in classification of HC and De, while Sumali et al. (2020) concluded that coefficients 1, 3, 4, 7, and 12 differentiate between depression and De.

While motor skills and linguistic semantic measurements are both somewhat good indicators in CI classifications, employment of acoustic phonetic measurements would make classification and diagnosis more time effective and cost friendly, easily done over the phone or on an application. Analysis on the other side would not require much human intervention as would semantic and syntactic analysis. It can be an all-automated process. Beyond CI classification, acoustic phonetic-related speech signal variables are also gaining traction for diagnosing other psychological conditions such as depression, stress, and bulimia (Hecker et al., 2022). This further underscores the potential future role of standardized acoustic phonetic variable sets as speech acoustics integrates with medicine and automatic disease detection. The adoption of standardized acoustic variable sets would facilitate a more accurate evaluation of machine learning models and their efficiency in disease classifications, as well as the development of algorithmic systems to meet the needs of efficient automated diagnosis.

While our data is confined to publications from the last 10 years, the study reveals a notable concentration of literature in the most recent years. This observation implies that more and more literature about this topic and related new trends would be published in the near future, which this study cannot capture. Furthermore, the data of the research analyzed in this review should not be generalized, given that many studies have relied on limited data for analysis and classification, potentially not fully representing the broader population. However, efforts are being made to compile corpora of patients' speech data (e.g., dementia bank) in different languages, which will address this limitation. This, along with the development and use of standard acoustic variable sets, would make cross-study comparisons more accurate and efficient. It is also worth mentioning that the field of machine learning is continuing to improve and develop, indicating that future studies' classification models may not rely on the models discussed in this review but on more improved and accurate models.

Lastly, studies that employ acoustic phonetic variables to

assess cognitive function lay the groundwork for transforming the evaluation process into an automated one that can be conducted in everyday settings as well as clinical environments. These studies foreshadow the potential for a cost-effective, non-surgical early detection and intervention approach for CI through automated cognitive function assessment. As a result, this study aims to emphasize the necessity for the development of a platform suitable for efficiently automating the CI assessment process using acoustic variables or speech signals.

In conclusion, it can be anticipated that the growing demand for standardized common grounds in the utilization of acoustic phonetic variables, both in classifying CI and other medical-related classifications, will lead to an increased use of standardized sets like ComPare, eGeMAPS, and the emergence of similar sets. And while the consideration of acoustic variables in CI classification is a relatively recent research trend, it has gained significant popularity due to the cost-efficiency of using speech signals as diagnostic determinants.

Reference

Agbavor, F., & Liang, H. (2022a). Artificial intelligence-enabled end-to-end detection and assessment of Alzheimer's disease using voice. *Brain Sciences, 13*(1), 28. doi:10.3390/brainsci13010028

Agbavor, F., & Liang, H. (2022b). Predicting dementia from spontaneous speech using large language models. *PLOS Digital Health, 1*(12), e0000168. doi:10.1371/journal.pdig.0000168

Beltrami, D., Gagliardi, G., Rossini Favretti, R., Ghidoni, E., Tamburini, F., & Calzà, L. (2018). Speech analysis by natural language processing techniques: A possible tool for very early detection of cognitive decline? *Frontiers in Aging Neuroscience, 10,* 369. doi:10.3389/fnagi.2018.00369

Boschi, V., Catricalà, E., Consonni, M., Chesi, C., Moro, A., & Cappa, S. F. (2017). Connected speech in neurodegenerative language disorders: A review. *Frontiers in Psychology, 8,* 269. doi:10.3389/fpsyg.2017.00269

Da Cunha, E., Plonka, A., Arslan, S., Mouton, A., Meyer, T., Robert, P., . . . Gros, A. (2022). Logogenic primary progressive aphasia or alzheimer disease: Contribution of acoustic markers in early differential diagnosis. *Life, 12*(7), 933. doi:10.3390/life12070933

de la Fuente Garcia, S., Ritchie, C. W., & Luz, S. (2020). Artificial intelligence, speech, and language processing approaches to monitoring alzheimer's disease: A systematic review. *Journal of Alzheimer's Disease, 78*(4), 1547-1574. doi:10.3233/JAD-200888

De Looze, C., Dehsarvi, A., Crosby, L., Vourdanou, A., Coen, R. F., Lawlor, B. A., & Reilly, R. B. (2021). Cognitive and structural correlates of conversational speech timing in mild cognitive impairment and mild-to-moderate Alzheimer's disease: Relevance for early detection approaches. *Frontiers in Aging Neuroscience, 13,* 637404. doi:10.3389/fnagi.2021.637404

De Stefano, A., Di Giovanni, P., Kulamarva, G., Di Fonzo, F., Massaro, T., Contini, A., . . . Cazzato, C. (2021). Changes in speech range profile are associated with cognitive impairment. *Dementia and Neurocognitive Disorders, 20*(4), 89-98. doi:10.12779/dnd.2021.20.4.89

Gonzalez-Moreira, E., Torres-Boza, D., Kairuz, H. A., Ferrer, C., Garcia-Zamora, M., Espinoza-Cuadros, F., & Hernandez-Gómez, L. A. (2015). Automatic prosodic analysis to identify mild dementia. *BioMed Research International, 2015,* 916356. doi:10.1155/2015/916356

Hall, A. O., Shinkawa, K., Kosugi, A., Takase, T., Kobayashi, M., Nishimura, M., . . . Yamada, Y. (2019). Using tablet-based assessment to characterize speech for individuals with dementia and mild cognitive impairment: Preliminary results. *AMIA Joint Summits on Translational Science Proceedings, 2019,* 34-43.

Hecker, P., Steckhan, N., Eyben, F., Schuller, B. W., & Arnrich, B. (2022). Voice analysis for neurological disorder recognition: A systematic review and perspective on emerging trends. *Frontiers in Digital Health, 4,* 842301. doi:10.3389/fdgth.2022.842301

Javeed, A., Dallora, A. L., Berglund, J. S., Ali, A., Ali, L., & Anderberg, P. (2023). Machine learning for dementia prediction: A systematic review and future research directions. *Journal of Medical Systems, 47*(1), 17. doi:10.1007/s10916-023-01906-7

König, A., Satt, A., Sorin, A., Hoory, R., Toledo-Ronen, O., Derreumaux, A., . . . David, R. (2015). Automatic speech analysis for the assessment of patients with predementia and alzheimer's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 1*(1), 112-124. doi:10.1016/j.dadm.2014.11.012

Kumar, M. R., Vekkot, S., Lalitha, S., Gupta, D., Govindraj, V. J., Shaukat, K., . . . Zakariah, M. (2022). Dementia detection from speech using machine learning and deep learning architectures. *Sensors, 22,* 9311. doi:10.3390/s22239311

Lin, H., Karjadi, C., Ang, T. F. A., Prajakta, J., McManus, C., Alhanai, T. W., . . . Au, R. (2020). Identification of digital voice biomarkers for cognitive health. *Exploration of Medicine, 1,* 406-417. doi:10.37349/emed.2020.00028

Liu, J., Fu, F., Li, L., Yu, J., Zhong, D., Zhu, S., . . . Li, J. (2023). Efficient pause extraction and encode strategy for Alzheimer's disease detection using only acoustic features from spontaneous speech. *Brain Sciences, 13*(3), 477. doi:10.3390/brainsci13030477

López-de-Ipiña, K., Alonso, J.-B., Travieso, C. M., Solé-Casals, J., Egiraun, H., Faundez-Zanuy, M., . . . Martinez de Lizardui, U. (2013). On the selection of non-invasive methods based on speech analysis oriented to automatic alzheimer disease diagnosis. *Sensors, 13,* 6730-6745. doi:10.3390/s130506730

Martínez-Nicolás, I., Llorente, T. E., Martínez-Sánchez, F., & Meilán, J. J. G. (2021). 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, 12,* 620251. doi:10.3389/fpsyg.2021.620251

Martínez-Nicolás, I., Llorente, T. E., Ivanova, O., Martínez-Sánchez, F., & Meilán, J. J. G. (2022). Many changes in speech through aging are actually a consequence of cognitive changes. *International Journal of Environmental Research and Public Health, 19*(4), 2137. doi:10.3390/ijerph19042137

Meilán, J. J. G., Martínez-Sánchez, F., Martínez-Nicolás, I., Llorente, T. E., & Carro, J. (2020). Changes in the rhythm of speech difference between people with nondegenerative mild cognitive impairment and with preclinical dementia. *Behavioural Neurology, 2020,* 4683573. doi:10.1155/2020/4683573

Nagumo, R., Zhang, Y., Ogawa, Y., Hosokawa, M., Abe, K., Ukeda, T., . . . Shimada, H. (2020). Automatic detection of cognitive impairments through acoustic analysis of speech. *Current Alzheimer Research, 17*(1), 60-68. doi:10.2174/1567205017666200213094513

NICE (National Institute for Health and Care Excellence). (2020). Clinical Knowledge Summaries: Dementia. https://cks.nice.org.uk/topics/dementia/

Nishikawa, K., Akihiro, K., Hirakawa, R., Kawano, H., & Nakatoh, Y. (2022). Machine learning model for discrimination of mild dementia patients using acoustic features. *Cognitive Robotics, 2,* 21-29. doi:10.1016/j.cogr.2021.12.003

Öhman, F., Hassenstab, J., Berron, D., Schöll, M., & Papp, K. V. (2021). Current advances in digital cognitive assessment for preclinical Alzheimer's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 13*(1), e12217. doi:10.1002/dad2.12217

Petti, U., Baker, S., & Korhonen, A. (2020). A systematic literature review of automatic Alzheimer's disease detection from speech and language. *Journal of the American Medical Informatics Association, 27*(11), 1784-1797. doi:10.1093/jamia/ocaa174

Riley, C. O., McKinstry, B., & Fairhurst, K. (2022). Accuracy of telephone screening tools to identify dementia patients remotely: Systematic review. *JRSM Open, 13*(9). doi:10.1177/20542704221115956

Shimoda, A., Li, Y., Hayashi, H., & Kondo, N. (2021). Dementia risks identified by vocal features via telephone conversations: A novel machine learning prediction model. *PLoS ONE, 16*(7), e0253988. doi:10.1371/journal.pone.0253988

Sumali, B., Mitsukura, Y., Liang, K., Yoshimura, M., Kitazawa, M., Takamiya, A., . . . Kishimoto, T. (2020). Speech quality feature analysis for classification of depression and dementia patients. *Sensors, 20,* 3599. doi:10.3390/s20123599

Tanaka, H., Adachi, H., Ukita, N., Ikeda, M., Kazui, H., Kudo, T., & Nakamura, S. (2017). Detecting dementia through interactive computer avatars. *IEEE Journal of Translational Engineering in Health and Medicine, 5,* 2200111. doi:10.1109/JTEHM.2017.2752152

Tang, F., Chen, J., Dodge, H. H., & Zhou, J. (2022). The joint effects of acoustic and linguistic markers for early identification of mild cognitive impairment. *Frontiers in Digital Health, 3,* 702772. doi:10.3389/fdgth.2021.702772

Themistocleous, C., Eckerström, M., & Kokkinakis, D. (2018). Identification of mild cognitive impairment from speech in swedish using deep sequential neural networks. *Frontiers in Neurology, 9,* 975. doi:10.3389/fneur.2018.00975

Themistocleous, C., Eckerström, M., & Kokkinakis, D. (2020). Voice quality and speech fluency distinguish individuals with mild cognitive impairment from healthy controls. *PLoS ONE, 15*(7), e0236009. doi:10.1371/journal.pone.0236009

Toth, L., Hoffmann, I., Gosztolya, G., Vincze, V., Szatloczki, G., Banreti, Z., . . . Kalman, J. (2018). A speech recognition-based solution for the automatic detection of mild cognitive impairment from spontaneous speech. *Current Alzheimer Research, 15*(2), 130-138. doi:10.2174/1567205014666171121114930

Vigo, I., Coelho, L., & Reis, S. (2022). Speech- and language-based classification of Alzheimer's disease: A systematic review. *Bioengineering, 9*(1), 27. doi:10.3390/bioengineering9010027

Voleti, R., Liss, J. M., & Berisha, V. (2020). A review of automated speech and language features for assessment of cognitive and thought disorders. *IEEE Journal of Selected Topics in Signal Processing, 14*(2), 282-298. doi:10.1109/jstsp.2019.2952087

Wang, H.-L., Tang, R., Ren, R.-J., Dammer, E. B., Guo, Q.-H., Peng, G.-P., . . . Wang, G. (2022). Speech silence character as a diagnostic biomarker of early cognitive decline and its functional mechanism: A multicenter cross-sectional cohort study. *BMC Medicine, 20,* 380. doi:10.1186/s12916-022-02584-x

WHO (World Health Organization). (2017). Global action plan on the public health response to dementia 2017-2025. Retrieved from http://apps.who.int/iris/bitstream/handle/10665/259615/9789241513487-eng.pdf;jsessionid=D684D2A563E5DD2D50C71D1E36C570CE?sequence=1

WHO (World Health Organization). (2023). Dementia. Retrieved from https://www.who.int/news-room/fact-sheets/detail/dementia

Yamada, Y., Shinkawa, K., Kobayashi, M., Nishimura, M., Nemoto, M., Tsukada, E., . . . Arai, T. (2021). Tablet-based automatic assessment for early detection of Alzheimer's disease using speech responses to daily life questions. *Frontiers in Digital Health, 3,* 653904. doi:10.3389/fdgth.2021.653904

Yeung, A., Iaboni, A., Rochon, E., Lavoie, M., Santiago, C., Yancheva, M., . . . Mostafa, F. (2021). Correlating natural language processing and automated speech analysis with clinician assessment to quantify speech-language changes in mild cognitive impairment and Alzheimer's dementia. *Alzheimer's Research & Therapy, 13,* 109. doi:10.1186/s13195-021-00848-x

Yoshii, K., Kimura, D., Kosugi, A., Shinkawa, K., Takase, T., Kobayashi, M., . . . Nishimura, M. (2023). Screening of mild cognitive impairment through conversations with humanoid robots: Exploratory pilot study. *JMIR Formative Research, 7,* e42792. doi:10.2196/42792

Appendix 1. Temporal and acoustic measures used in assessment in reviewed articles

Study	Temporal and acoustic measures used in assessment		
Toth et al. (2018)	1. Hesitation ratio: More than 30 ms no speech 2. Speech tempo	3. Length and number of silent and filled pauses 4. Length of utterance	
Nagumo et al. (2020)	1. Mean and *SD*s of F1 and F2 2. Mean and *SD*s of F0 3. Triangular vowel space area 4. Vowel articulation index	5. Utterance duration 6. Number of pauses 7. Length of pauses 8. Number of uttered syllables	9. Mean and *SD* of syllable duration 10. Mean and *SD* of the power of maximum and minimum
Gonzalez-Moreira et al. (2015)	1. Speech time 2. Number of pauses 3. Proportion of pause 4. Phonation time	5. Proportion of phonation 6. Speech rate 7. Articulation rate 8. Number of syllables	9. Mean of syllables duration 10. *SD* of F0 11. Maximum variation of F0 12. Mean of F0
König et al. (2015)	1. Mean median *SD* and ratio mean and sum of duration 2. Voice segment length 3. Silence segment length	4. Periodic segment length 5. Aperiodic segment length 6. Vocal reaction time 7. Length of sentence duration	8. Amount of silence 9. Distance time of the second, third, fourth until the ninth detected word position from the first word position.
De Stefano et al. (2021)	1. Number of semitones 2. Sound pressure level maximum (SPLmax)	3. Sound pressure level medium 4. Speech time	
Meilán et al. (2020)	1. Reading time 2. Phonation time 3. Speech rate 4. Pauses number 5. Normalized_PVI 6. Syll_Interv_DAverage 7. Interv_*Δ*Standar 8. VARCO of syllabic interval duration	9. Syllabus interval number 10. Articulation rate 11. F0 (Mean, asymmetry (skewness), center of gravity and center of gravity *SD*) 12. LTAS (Mean, *SD*, range, 1 kHz-2 kHz, 2 kHz-4 kHz)	13. Intensity (Mean, mean_*SD*, amplitude minimum, intens diferenc max-min mean) 14. Jitter (Local) 15. Shimmer (Local) 16. Voice breaks 17. Acoustic Voice Quality Index (AVQI) 18. HNR (dB)
Kumar et al. (2022)	1. Jitter 2. Shimmer 3. F0	4. (First four) formants 5. MFCC 6. GTCC	7. Delta-MFCC and delta-GTCC
Shimoda et al. (2021)	1. Start and end time of all sounding and silent intervals 2. Intensity	3. Pitch 4. Center of gravity, skewness, kurtosis, and *SD*	
Liu et al. (2023)	1. ComParE: Energy, spectral, MFCC, HNR, voice quality features, viterbi smoothing for F0, spectral harmonicity, and psychoacoustic spectral sharpness.	2. eGeMAPS: F0 semitone, jitter, shimmer, loudness, spectralflux, MFCC, F1, F2, F3, alpha ratio, Hammarberg index, and slope V0 features.	
Lin et al. (2020)	1. JitterDDP_sma_de 2. Mfcc_sma_de[4] 3. ShimmerLocal_sma_de 4. Mfcc_sma_de[3] 5. Pcm_zcr_sma_de	6. Mfcc_sma_de [1] 7. Pcm_RMSenergy_sma 8. JitterLocal_sma 9. Adspec_lengthL1norm_sma 10. JitterDDP_sma	11. VoicingFinalUnclipped_sma 12. F0final_sma_de 13. F0final_sma 14. AudspecRasta_lengthL1norm_sma
Themistocleous et al. (2018)	1. Vowel formants: First five formant frequencies, total duration 2. F0 (Mean, min, and max)	3. Vowel duration: From vowel onset to vowel offset.	
Da Cunha et al. (2022)	1. Vocal reaction time 2. Vowel phonation time 3. Consonant phonation time 4. Phonation time deviation	5. Pause ratio 6. Non-silent pause ratio 7. Silent pause ratio 8. Speech rate	9. Intensity range 10. F0: Minimum, maximum
Nishikawa et al. (2022)	1. Vowel sounds (/a/, /i/, /u/, /e/, /o/) 2. MFCC	3. F0 4. Formant frequency (F1, F2)	5. Jitter (Local jitter, PPQ5) 6. Shimmer 7. HNR
Martínez-Nicolás et al. (2022)	1. Duration (Reading time) 2. Number of pauses speech rate 3. Average duration of syllabic intervals	5. *SD* of syllabic 6. Intervals duration mean amplitude 7. LTAS_50-1K 8. First formant (F1) *SD*	9. F0 10. Spectral skewness 11. HNR 12. Jitter (Local)
López-de-Ipiña et al. (2013)	1. Pitch (*SD*, max and min) 2. Intensity (*SD*, max and min) 3. Period (Mean, *SD*) 4. Root mean square amplitude	5. Shimmer 6. Local jitter 7. NHR 8. HNR	9. Autocorrelation. 10. Fraction of locally unvoiced frames 11. Degree of voice breaks.
Yoshii et al. (2023)	1. Jitter (Local, relative average perturbation [rap], Point period perturbation quotient [ppq5], difference of difference of periods [ddp])	2. Duration of response time 3. Duration of speech periods 4. F0cov (Coefficient of variation of the fundamental frequency)	5. Shimmer (Local, amplitude perturbation quotient [apq]3, apq5, apq11, average absolute differences between the amplitudes of consecutive periods [dda])
Sumali et al. (2020)	1. Pitch 2. HNR 3. Zero-crossing rate	4. MFCC 5. GTCC 6. Mean, median frequency	7. Signal energy 8. Spectral centroid 9. Spectral roll off point
Wang et al. (2022)	1. Silent periods		
Yamada et al. (2021)	1. MFCC (Skewness and kurtosis) 2. First three formant frequencies 3. Jitter (Local, RAP, PPQ5, DDP) 4. Shimmer (Local, APQ3, APQ5, APQ11, DDA)	5. Pitch and its variation (Median absolute deviation) 6. Speech rate 7. Phoneme rate 8. Phonation time	9. Pause 10. Response time (Mean, median, *SD*, maximum, and minimum, of each feature)
Hall et al. (2019)	1. Means and *SD*s of length for speech and silent segments. 2. F0 (Mean and *SD*)	3. F1-F5 (Means and *SD*) 4. Jitter 5. Shimmer (Local and APQ3)	
Themistocleous et al. (2020)	1. H1-H2, H1-A1, H1-A3 2. Cepstral peak prominence 3. Mean energy concentration 4. Hammarberg Index	5. Shimmer 6. Jitter 7. Harmonicity 8. Articulation rate	9. Average syllable and articulation rate 10. Speech rate

Note. F0=fundamental frequency; GTCC=gammatone cepstral coefficients; HNR=harmonic-to-noise ratio; MFCC=mel-frequency cepstral coefficients; NHR=noise-to-harmonic ratio; SRP=speech range profile; *SD*=standard deviation.