Journal of Alzheimer's Disease 100 (2024) 1–27
DOI 10.3233/JAD-231459
IOS Press

Systematic Review

Machine Learning Approaches for Dementia Detection Through Speech and Gait Analysis: A Systematic Literature Review

Mustafa Al-Hammadi[a,*], Hasan Fleyeh[a], Anna Cristina Åberg[b,c], Kjartan Halvorsen[b] and Ilias Thomas[a]
[a]*School of Information and Engineering, Dalarna University, Falun, Sweden*
[b]*School of Health and Welfare, Dalarna University, Falun, Sweden*
[c]*Department of Public Health and Caring Sciences, Geriatrics, Uppsala University, Uppsala, Sweden*

Accepted 19 April 2024
Pre-press 3 June 2024

Abstract.
Background: Dementia is a general term for several progressive neurodegenerative disorders including Alzheimer's disease. Timely and accurate detection is crucial for early intervention. Advancements in artificial intelligence present significant potential for using machine learning to aid in early detection.
Objective: Summarize the state-of-the-art machine learning-based approaches for dementia prediction, focusing on non-invasive methods, as the burden on the patients is lower. Specifically, the analysis of gait and speech performance can offer insights into cognitive health through clinically cost-effective screening methods.
Methods: A systematic literature review was conducted following the PRISMA protocol (Preferred Reporting Items for Systematic Reviews and Meta-Analyses). The search was performed on three electronic databases (Scopus, Web of Science, and PubMed) to identify the relevant studies published between 2017 to 2022. A total of 40 papers were selected for review.
Results: The most common machine learning methods employed were support vector machine followed by deep learning. Studies suggested the use of multimodal approaches as they can provide comprehensive and better prediction performance. Deep learning application in gait studies is still in the early stages as few studies have applied it. Moreover, including features of whole body movement contribute to better classification accuracy. Regarding speech studies, the combination of different parameters (acoustic, linguistic, cognitive testing) produced better results.
Conclusions: The review highlights the potential of machine learning, particularly non-invasive approaches, in the early prediction of dementia. The comparable prediction accuracies of manual and automatic speech analysis indicate an imminent fully automated approach for dementia detection.

Keywords: Alzheimer's disease, cognitive impairment, deep learning, dementia disorders, gait analysis, machine learning, non-invasive, speech analysis

INTRODUCTION

Worldwide, dementia is a global concern as the aging population continues to grow. By 2050 the number of people over 60 years old is expected

*Correspondence to: Mustafa AL-Hammadi, School of Information and Engineering, Dalarna University, 79131 Falun, Sweden. Tel.: +46 766351237; E-mail: mum@du.se.

to double to 2.1 billion. One of the most common health issues with aging is neurodegenerative disorders; however, it is crucial to recognize that not all forms of dementias are neurodegenerative [1]. According to the World Health Organization (WHO), there are currently 55.2 million people living with dementia, which makes it the 7th leading cause of death worldwide. The estimated global societal cost of dementia in 2019 was estimated to US$ 1.3 trillion [2]. Dementia is an umbrella term under which the most common form is Alzheimer's disease (AD), accounting for approximately 60% to 80% of cases. Dementia progressively degrades the brain neurons over a period of twenty years or even more, often without any noticeable symptoms. This period of undetectable progression makes it challenging for clinicians to identify the condition in its early stages [3]. As the disease progresses, the ability for activities of daily living (ADL) declines in people with dementia, and as a result, they become dependent on their families and carers.

As there are currently no disease-modifying therapies for dementia, early detection is crucial for early intervention, planning, and assistance. Early detection can mitigate the risk of the disease progressing from mild cognitive impairment (MCI) into a severe condition of dementia. MCI can be a prodromal state of dementia, which does not interfere significantly with performing ADL [4, 5].

There are different methods for assessing cognitive impairments and discriminating among various forms of dementia. On the one hand, clinical cognitive tests evaluate memory, attention, and language skills but require healthcare expertise. The Mini-Mental State Examination (MMSE) and the Montreal Cognitive Assessment (MoCA) are crucial in evaluating cognitive function and helping in neuropsychological diagnostic differentiation within dementia detection. The MMSE efficiently screens for cognitive impairment, focusing on orientation, memory, and language, while the MoCA provides a comprehensive analysis, sensitive to mild cognitive impairment and early dementia stages, with an emphasis on executive functions and visuospatial abilities [6]. Brain imaging is costly and time-consuming. Other methods such as cerebrospinal fluid (CSF) analysis, electroencephalography (EEG), and blood tests are invasive methods. On the other hand, marker-free video recordings of the gait or speech performance are emerging as non-invasive methods for dementia detection [7]. Through video recording, movement can be recorded under a single or dual-task paradigm.

Dual-task is a test that refers to the process of walking while performing cognitive or motor tasks simultaneously [8]. Time-up-and-go test (TUG), and time-up-and-go dual-task (TUGdt) are two examples of movement tasks. The TUG test is employed to evaluate a person's mobility and balance [9]. TUG would usually entail the subjects standing up from a chair, walking a distance, and returning to the chair, while TUGdt would engage individuals in simultaneous cognitive activities, such as walking while performing a memory task such as counting backward, naming animals, objects, or reciting a list. Similarly, various speech tasks can be employed to assess individuals for signs of dementia. These tasks encompass spontaneous speech, picture description, verbal fluency, and reading aloud are examples of speech tasks that can be used to assess dementia detection [10]. Extracted features from gait and speech data can serve as non-invasive biomarkers for early dementia detection.

Motor performance and cognition are closely interlinked with each other. Gait analysis can provide pivotal information to distinguish cognitive impairments and has the potential to be an early marker for dementia [11]. Gait refers to the pattern of walking and encompasses spatiotemporal parameters, for instance, walking speed (4-meters gait speed ($<$0.8 m/s) is impaired, (0.8 to $<$ 1 m/s) moderate, and intact ($\geq$1 m/s) [12], stride length, and gait variability. Motor impairments and subtle alterations in gait patterns can serve as potential indicators of cognitive decline, including the development of dementia [13, 14]. Slower gait speed along with increased gait variability and difficulties in performing dual-task can manifest early before MCI occurs [15, 16]. Alteration in gait could be determined by observing changes in walking patterns during dual-task activities, such as performing verbal fluency tasks, which can exacerbate gait irregularities and indicate cognitive-motor interference. Dual-task can be a valuable tool in the early detection and assessment of dementia and offers a comprehensive window into an individual's cognitive functioning [17]. It plays an important role in the early detection and holistic assessment of dementia, granting clinicians and researchers the opportunity to gain insights into their cognitive functioning.

Speech impairment could be another symptom of cognitive decline. Speech data can be used for the non-invasive extraction of digital biomarkers. Speech analysis provides a cost-effective, simple, and easily accessible means of screening individuals for dementia. In addition, speech data can be collected in large

volumes. In the early-stage subjects with dementia often experience difficulties in finding the right word, forming coherent sentences, loss of verbal fluency, slow and hesitant speech [18, 19]. Moreover, linguistic features (i.e., part of speech, word count, and word embeddings), and acoustic features (i.e., spectral features, pauses, and speech rhythm) in spontaneous speech can serve as features for identifying cognitive decline [20, 21].

Machine learning is already widely used in healthcare domains such as neurodegenerative diseases. By analyzing substantial datasets that contain diverse features, it has the capability to identify subtle patterns that could go unnoticeable by clinicians and aid in the diagnosis of diseases, including dementia. This results in the improvement of early detection of different forms of cognitive impairments and has the potential to enhance the quality of life for individuals affected by the disease, enabling early diagnosis, timely intervention, and well-informed medical decisions. The focus of this study is on gait or speech analysis because they are non-invasive and cost-effective diagnostic tools for dementia. Furthermore, their combination has been underexamined in clinical studies so far. As the field continues to advance, the aim of this systematic literature review is to summarize the existing literature on machine learning methods used for dementia prediction through gait or speech analysis and pinpoint the best practices. It addresses the following research questions: Which machine learning algorithms were utilized in the studies reviewed, and which among them have proven to be the most effective? Additionally, this literature review examines the most important features in the included studies. Finally, analyze the characteristics of the data used for the selected gait studies and the characteristics and languages of the data used for the selected speech studies. By synthesizing existing knowledge, this review strives to inform best practices and steer future research efforts in this critical field.

Previous systematic literature review papers in the field of dementia detection through gait and speech analysis have provided valuable insight and made significant contributions. However, the reviews that are based on gait analysis are either focused on employing simple statistical methods without machine learning [17, 15, 22] or sensor-based methods [23–25]. To our knowledge, no systematic literature review paper has been conducted on the application of machine learning for dementia prediction through the analysis of gait by non-invasive methods. Additionally, prior systematic review papers on speech have examined the literature up until 2019 and 2020 [26–28] or focused exclusively on deep learning methods in the context of dementia detection from speech [29]. However, as the field is continuously evolving, there is a need for a more up-to-date summary of the current state-of-art methods, particularly examining all machine learning methods, and not focusing only on one aspect. In the literature, no review has examined the fusion of gait and speech, to examine if a multimodal evaluation of dementia would improve early diagnosis, evaluating the potential of TUGdt.

The main contribution of this review is to address the aforementioned gaps and offer a comprehensive synthesis of the best practices of machine learning approaches for dementia detection using speech or gait data. The features employed for classification in machine learning models, characteristics of the data, and languages utilized in the speech analysis are investigated. The findings of this study make a significant contribution to the field by highlighting promising yet underexplored methods, thereby guiding the direction of further research and innovation.

METHODOLOGY

A search was conducted using The PRISMA protocol [30]. This type of review is known for its methodical, well-established, and rigorous approach, which is crucial for ensuring the transparency and reproducibility of the results [31]. A systemic literature review is conducted for the purpose of comprehensively summarizing the existing literature on a particular topic, while also finding gaps in the body of knowledge that point to potential venues for new research.

The process of a systematic literature review involves various key stages, including formulating the research question, searching multiple databases, screening, and selecting relevant studies, and finally extracting data and reporting the findings [32]. PRISMA also emphasizes the importance of conducting meta-analysis when appropriate; however, meta-analysis is not done in this case due to heterogeneity in the data of the included studies.

Search strategy

Comprehensive searches were conducted in three selected databases (Scopus, PubMed, Web of Science) to identify relevant studies published from 2017 to 2022. This period captures the recent advancement in machine learning techniques with the highest

accuracies and the latest tools and methodologies for increased relevance in both current and future research. The following search term was used through an iterative process of refining the search keywords to achieve comprehensive results: (Dementia OR Alzheimer OR "cognitive impairment") AND (TUG OR "dual task" OR gait OR face OR "facial recognition" OR speech OR voice) AND (learning OR predict OR "artificial intelligence"). Table 1 shows the search keywords used in the Scopus database:

Table 1
Search string used in the Scopus database showing keywords along with the date of accessing the databases

Search date	January 13, 2023
Search query	
TITLE-ABS-KEY ((dementia OR alzheimer OR "cognitive impairment") AND (tug OR "dual task" OR gait OR face OR "facial recognition" OR speech OR voice) AND (learning OR predict OR "artificial intelligence")) AND (LIMIT-TO (PUBYEAR, 2022) OR LIMIT-TO (PUBYEAR, 2021) OR LIMIT-TO (PUBYEAR, 2020) OR LIMIT-TO (PUBYEAR, 2019) OR LIMIT-TO (PUBYEAR, 2018) OR LIMIT-TO (PUBYEAR, 2017)) AND (LIMIT-TO (LANGUAGE, "English"))	

The search in Web of Science was conducted using the 'topic' filter which includes the title and the abstract. The PubMed search was conducted using 'all filed' filter which includes title, abstract, and full text. In Scopus, the search was conducted in the abstract and title. The search keywords were intended to be comprehensive and cover all the literature related to dementia detection using machine learning through gait and speech analysis. The databases were accessed on 13 January 2023.

The PICOS table is a structured framework employed in systematic literature reviews to define the research question and guide the search strategy by outlining the Population, Intervention, Comparison, Outcome, and Study Design. This approach ensures a comprehensive and focused search, enabling the identification and selection of relevant studies for inclusion in the review [33]. The PICOS table can be seen in Table 2.

Table 2
PICOS table for the search strategy in dementia prediction using machine learning through non-invasive methods

Component	Description
Population (P)	Studies related to dementia and its subtypes. The population includes diverse age groups, genders, and stages of cognitive decline, targeting those who are subjects of non-invasive diagnosis tools.
Intervention (I)	Not applicable.
Comparison (C)	Not applicable.
Outcomes (O)	Identification of machine learning algorithms used for dementia classification.
Study Design (S)	The review encompasses a variety of study designs such, as cohort studies, longitudinal studies, prospective cohort studies, and cross-sectional studies, with a particular focus on those employing non-invasive methodologies to ensure a comprehensive synthesis of relevant evidence.

Inclusion criteria

In this systematic literature review, the search was not limited to specific authors, subject areas, or document types. Instead, it was conducted by a comprehensive search across the aforementioned databases, including journals, conference proceedings, and book chapters. Through this inclusive search, the study aims to reduce the potential for publication bias. Furthermore, it included all relevant literature, both paid and subscription-based papers. This approach reduced the risk of not including valuable studies and ensures a thorough representation of the available evidence. The only restrictions that were placed were on studies that were not written in English and publications that were not published within the timeframe of 2017 to 2022. By including a wide range of sources and avoiding any narrow focus, the aim was to conduct a thorough and unbiased review of the available literature on the research topic.

The Inclusion criteria in this review were: 1) Studies written in languages other than English were excluded; 2) Studies between 2017 to 2022; 3) Studies that utilized machine learning approaches for dementia detection or studies that employ speech analysis for dementia detection; 4) Studies using gait analysis for dementia detection; 5) If speech data is automatic or both automatic and manual transcription of the speech data; 6) All types of papers (journals, conference proceedings, and book chapters). The exclusion criteria were: 1) Sensor-based methods of data collection; 2) Invasive studies for instance, MRI, PET, and EEG; 3) Simple statistics methods; 4) Examining only association without prediction; 5) Studies that focus on other neurodegenerative diseases; 6) Short conference contributions (<4 pages); 7) Public databases (DementiaBank, ADReSS Challenge dataset, and Framingham Heart Study (FHS)) and the reason is that these studies use the same dataset and same methods.

Search outcome

In total 3,819 studies were identified from the selected electronic databases as follows: Web of Science total was 1,197, Scopus total was 1,285, PubMed total was 1,337 articles. After removing the duplicates, the total was 2,655. After screening the title and abstract ($n = 300$) were selected for full-text evaluation. To reduce the bias of the selected studies, full-text papers ($n = 300$ studies) were reviewed independently by two reviewers (MA and IT) and only the papers were both reviewers selected as relevant were included. The level of agreement between the two reviewers was 90%. The reviewers discussed any disagreements and finalized the papers to be included based on the consensus. Incorporating this method enhances the objectivity and reliability of the data extraction and synthesis processes and avoids individual biases in interpreting and synthesizing the data. In total 40 publications were included, 9 for gait and 31 for speech, the distribution of the publications per year can be seen in Fig. 1. The flowchart of the methodology can be seen in Fig. 2. This figure compiles the selection process according to search

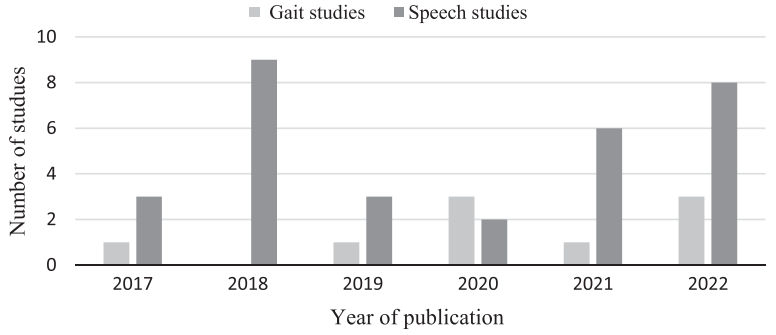

Fig. 1. Number of publication per year for gait and speech studies included for synthesis.

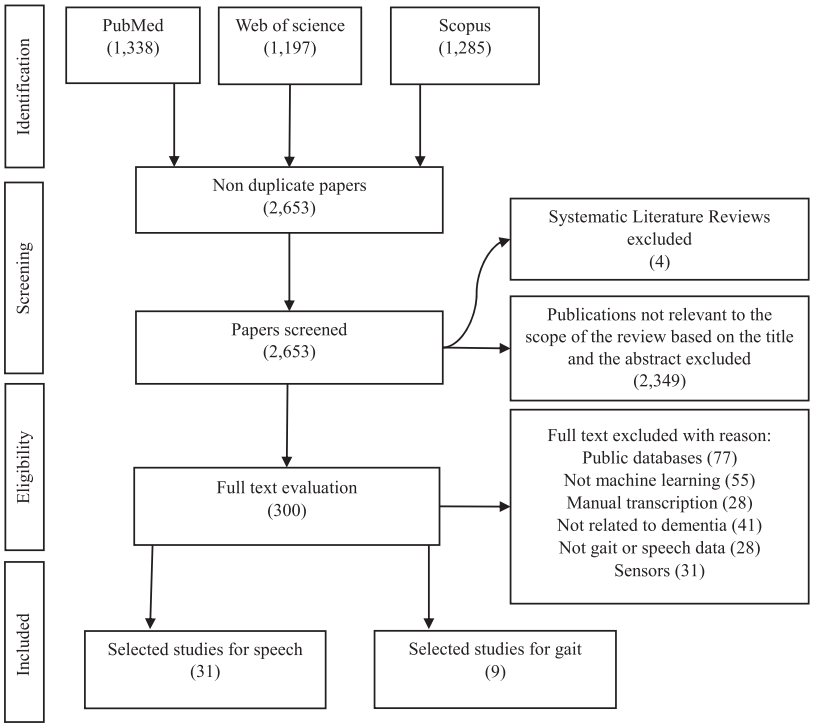

Fig. 2. Methodology flowchart following the PRISMA guidelines. Stages of literature selection starts by databases identification, reports screening, full text evaluation, and finally 40 papers were selected for synthesis according to the inclusion/exclusion criteria. The Figure shows the count of excluded papers with reason.

methods and PRISMA flow diagram for systematic reviews [30]. After inclusion/exclusion criteria and duplicates removal, 300 reports were sought for retrieval, and after assessing eligibility, 40 studies were included in this review.

RESULTS

This section is organized into two sections for gait and speech analysis. Within each section, the selected studies are comprehensively synthesized separately. By presenting the synthesized literature in this manner, the aim was to provide a clear and structured way of the key themes and trends coming from the literature in each domain.

Two tables have been created, one for speech and one for gait. Table 3 for gait, consists of the author's name, publication year, study design, dataset, demographic characteristics, instrument, algorithms used, input variable, output variable, and best accuracy.

Table 5 created for speech comprises of author's name, publication year, speech task, language of the speech task, dataset, demographic characteristics, features, algorithms used, validation methods, and best accuracy.

Gait analysis for dementia detection using machine learning

Gait assessment for dementia detection through machine learning has the potential to be a non-invasive tool for early diagnosis. Gait, which is the pattern of walking, has the potential to provide valuable insight into cognitive health. Figure 3 illustrates the whole process of dementia detection using machine learning. Within this part, the study starts by illustrating the machine learning techniques presented in the included gait paper studies, and features used for classification.

Machine learning techniques

All the studies reviewed for gait analysis utilized supervised learning. The most frequent machine learning algorithm that is used in the selected studies is support vector machine (SVM) ($N = 5$), followed by logistic regression (LR) ($N = 3$), then decision trees (DT) ($N = 2$). Only three studies used deep learning algorithms (DP) for gait analysis [34–36]. The percentage of each algorithm is presented in Fig. 4.

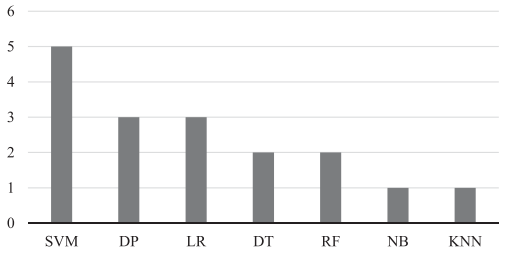

Fig. 4. Overview of the count of identified machine learning methods used in gait studies. SVM, support vector machine; DP, deep learning; LR, logistic regression; DT, decision trees; RF, Random Forest; NB, Naïve Bays; KNN, K-Nearest Neighbor.

Here is a brief description of the algorithms and how they have been used in the included studies:

Application of neural network

Neural network (NN) is a computational model inspired by the human brain. It encompasses of interconnected neurons that process the input data in a layered manner. Likewise, the human brain adjusts the synaptic weights between neurons, these architec-

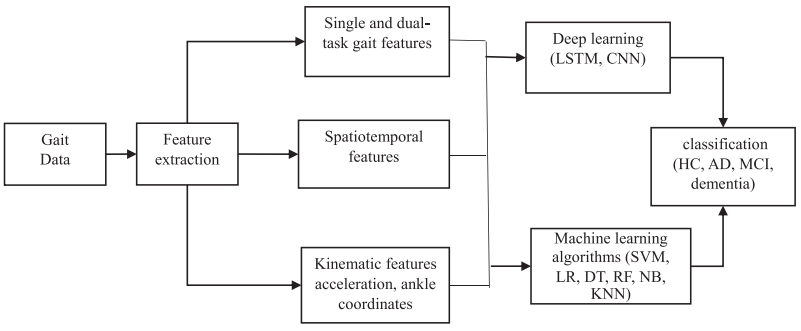

Fig. 3. Overview of identified machine learning methodology for dementia detection using gait analysis. HC, healthy controls; AD, Alzheimer's disease; MCI, mild cognitive impairment; SVM, support vector machine; LR, logistic regression; DT, decision trees; RF, Random Forest; NB, Naïve Bays; KNN, K-Nearest Neighbor.

tures adjust the weights to enable them to learn and improve their performance. NNs contain activation functions to decide if a neuron should be activated or not based on the weighted sum [37].

NN was implemented by Ishibashi et al. [34], for extracting and utilizing features from Doppler radar images for dementia classification. The feature-extraction of gait data is done through image binarization, region segmentation, and pixel counting. The study compared the performance of different models, including K-Nearest Neighbors (KNN), Naïve Bayes (NB), Random Forest (RF), Decision Trees (DT), and ensemble modeling (EN). The result showed that NN outperformed other classifiers, achieving an accuracy of 82.1% for the classification of dementia subjects from healthy controls (HC). However, in terms of recall, the ensemble based on DT, RF, KNN, and NN accomplished 72.2%. SVM and NB achieved a low recall score (41.7%), and the reason might be the imbalanced dataset with many more healthy controls than patients with dementia.

You et al. [35] aimed to explore gait parameters for the early diagnosis of MCI and AD by implementing the Long-Short-Term-Memory (LSTM) and RF algorithms. The authors explored whole-body movement to identify important parameters for dementia stratification from skeleton joints captured by a Kinect sensor. The system demonstrated an increase of 4.54% accuracy when using both upper and lower body part features compared to employing only lower body features. This study utilized gait speed, gait cycle, stride length variation, gait cycle variation, hand swing, and head posture variation as features for classification. The results showed that LSTM (accuracy of 90.48%, sensitivity of 92.00%, and specificity of 88.24%) performed better than RF.

Zhang et al. [36] presented a dementia detection system using gait analysis captured through a Kinect 2.0 camera. Features were extracted from human skeleton joints from both single task and dual-task in which subjects count down from 20 and subtract 3 each time. The features learned from a Convolutional Neural Network (CNN) outperformed handcrafted features such as mean speed, standard deviation of speed, mean step length, mean gait cycle time, and mean acceleration. In the context of positive case classification, sensitivity is important to evaluation metrics, and CNN attained the highest sensitivity of 74.10%. However, in terms of accuracy, SVM achieved a better result (60.42%)

Application of support vector machines

SVM is a supervised machine learning algorithm, it was developed in the 1990 s and has commonly been used for classification tasks. SVM makes use of kernels which are mathematical functions that transform the input data into a higher-dimensional space to allow for separation between data points [38]. SVMs were utilized in five studies and are the most used machine learning technique for dementia detection research among our search results.

Seifallahi et al. [39] developed an SVM classifier and out of 61 features extracted from the TUG test they identified 12 significant gait features for the classification of AD from HC after modifying for age and Geriatric Depression Scale. This study achieved an accuracy of 97.75% and an F-score of 97.67% for five-fold cross-validation and 98.68% and 98.67% for leave-one-subject-out cross-validation. This study provides a low-cost and convenient tool for AD detection through quantifying TUG tests in older adults. The tool has the capability to enable widespread AD diagnosis in clinical or non-clinical settings, which can help in the early detection of AD.

Kondragunta et al. [40] implemented an SVM to classify healthy individuals from mild or severe cognitive impairments. 3D-based pose estimation was implemented using depth data to quantify gait parameters and gait cycles of the elderly. The estimated gait parameters and cycles were used by Dynamic Time Wrapping to compare the patterns of the gait cycle of the subjects in various trials such as Regular Gait 1 (RG1), Regular Gait 2 (RG2), Counting Backward 1 (CB1), Counting Backward 3 (CB3), Fast Gait (FG) and Words with Special Letters (WSPL) and then fed to SVM for classification. The overall accuracy is low with the highest accuracy of 69% CHI versus pMCI versus MCI. The F1 score is best for CHI versus MCI with 80%. One suggested reason for the low classification accuracy is the features selected are not sufficient. For instance, the spatiotemporal and demographic features are not used in this study, and a similar pattern of gait features among the elderly over 80 years old.

Aoki et al. [41] applied an SVM classifier and extracted gait features from the whole-body movement. The optimal MMSE cut-off score of 25 achieved the highest AUC when participants split into either positive or negative. The study showed that dual-task gait features obtained better results as compared to single gait features with ROC 59.8%, and

74.7% respectively. Moreover, the study presented that the rich features demonstrate superiority in comparison to the simple features i.e., AUC (66.6%, 74.7%) respectively.

Application of logistic regression

LR is a machine learning method that is used for binary classification in order to model the probability of an event occurring based on input. It allows for the estimation of odds ratios, which quantify the impact of predictors on the likelihood of the event. The logistic transformation is implemented to map the input value between 0 and 1 [42].

Åhman et al. [8] used LR and showed that TUGdt enhances dementia prediction beyond sole reliance on demographic variables and cognitive test alone (c-statistics from 88% to 94%) and the single-task test (c-statistics from 86% to 89%) conducted separately. Notably, the improvement in the prediction was observed in the younger patients under the age of 72 years.

Nielsen et al. [43] employed the LR model to examine the ability of the dual-task to discriminate between MCI versus HC and MCI versus AD. The dual-task incorporated the TUG test combined with backward counting starting from 100 and decreasing by one,

Table 3
Summary of the selected gait studies

Author	Instrument	Study design	Dataset	Demographic characteristics	Algorithm	Input variable	Output variable	Best Accuracy
Ishibashi et al. [34] (2022)	Doppler radar	CS	EPDDRI Dataset images = 117	–	a) SVM, b) KNN, c) NB d) DT, e) RF, f) NN	Gait data	Dementia, HC	f) Acc = 82.1%
Seifallahi et al. [39] (2022)	Kinect V.2 camera	CS	N = 85 (HC = 47, AD = 38)	Female=46 (HC = 25, AD = 21), Male = 39, Mean age (72.11)	SVM	25 joint positions of the TUG test	HC, AD	Acc = 97.75%
You et al. [35] (2020)	8 Kinect V2 Cameras	CS	N = 88 (HC = 35, MCI or AD = 53)	Male=59 Female=29 Age>46	a) LSTM with rearrange data, b) RF of LB, c) RF with features	Gait features, all skeleton joint (upper and lower body features	HC, AD\MCI	a) Acc = 90.48%
Zhang et al. [36] (2021)	Kinect 2.0 camera	CS	N = 300 (HC = 106, Dementia = 194)	Female=200, Male = 100, Mean age = 84.5	a) CNN, b) SVM, c) DT	25 skeleton joints, Single Gait, DT	HC, Dementia	b) Acc = 58.75%
Åhman et al. [8] (2020)	Video camera, stopwatch	LS	N = 172 (SCI = 61, MCI = 111	Mean age = 71 Age range (39–91) Female=78, Male = 94	LR	TUGdt, cognitive test, demographic characteristics	Dementia Incidence	AUC-ROC = 94%
Kondra -gunta et al. [40] (2020)	Microsoft Kinect V2	LS	N = 142	age>80 years	DTW for features extractions, SVM	Gait cycle, 25 key joints, single-task, dual-task gait	HC, CI	CHI versus MCI ACC = 69%
Aoki et al. [41] (2019)	Kinect senor	CS	N = 102 (MMSE positive = 15, MMSE negative = 88)	Female=76, male = 26, age range of n = 63 is (68–90)	SVM	DT gait features, single gait features	HC, CI	ROC = 74.7%
Nielsen et al. [43] (2017)	Stopwatch	PCS	N = 86 (HC = 41, MCI=28, 17 = AD)	Mean age (68.66), Male = 60.033%, Female = 39.96%	LR	TUGdt test, DT Cost	HC, MCI, AD	MCI versus HC AUC of 82%, and MCI versus AD AUC of 73%
Ng et al. [44] (2022)	10 m Walkway, Stopwatch, MIC	PCS	N = 2,544	Age $\geq$ 55	LR	TUG test, GS, KES, POMA	MCI, CD	AUC = 72.9%

CS, cross-sectional; LS, longitudinal study; CI, cognitive impairment; CHI, cognitively healthy individuals; DT, dual-task; PCS, Prospective Cohort Study; GS, fast gait speed; TUG, Time Up And Go; KES, knee extension strength; POMA, Performance Oriented Mobility assessment; CD, cognitive decline; LB, lower body; SVM, support vector machine; KNN, K-nearest neighbor; NB, Naïve bias; DT, Decision tree; RF, Random forest; NN, Neural network; LSTM, long short-term memory; CNN, Convolution neural network; LR, Logistic regression; DTW, Dynamic time warping; HC, healthy control; AD, Alzheimer's disease; MCI, mild cognitive impairment; Acc, Accuracy; SCI, subjective cognitive impairment; MIC, maximum isometric contraction; AUC, Area under the [ROC] curve; ROC, receiver operating characteristic.

assessing the performance decline as dual-task cost. The study achieved an AUC of 78% for discriminating between MCI from the HCs, an AUC of 73% among MCI and AD, and an AUC of 90% between HC and AD through the TUGdt (the TUGdt has a high discriminating ability between MCI and HC and a moderate discriminating ability between MCI and AD).

Ng et al. [44] implemented LR and demonstrated the decrease in the TUG test of functional mobility is prodromal to the occurrence of MCI. TUG displayed the maximum accuracy (AUC = 72.9%) in predicting incidents of MCI-dementia, followed by the Gait Speed (GS) test (AUC = 68.3%), Knee Extension Strength (KES) test (AUC = 62.4%), and the Performance Oriented Mobility Assessment (POMA) test (AUC = 56.1%). POMA is a functional assessment tool utilized to evaluate balance, gait, and mobility in elderly individuals, providing a score that indicates their risk of falls [45].

Study characteristics

Of the nine papers selected for the detection of dementia from gait, three focused on MCI, one on AD, two on both AD and MCI, and three on the broader aspect of dementia classification. In terms of gait data, two studies used the TUG test, two studies used the normal gait test, two studies performed the TUGdt, and three studies incorporated both single and dual-task. Regarding the method of collecting the gait data, five studies utilized Kinect sensors, three studies employed stopwatches, one study used video cameras, and one study used maximum isometric contraction. Five studies used cross-sectional as the study design method, two longitudinal designs, and two prospective cohort study designs. The extracted data is shown in Table 2.

Feature characteristics

Several studies have utilized diverse features for gait analysis as a tool for the early detection of dementia. Seifallahi et al. [39] identified numerous features related to sit-to-stand, turning, and stand-to-sit subtasks and important features chosen from different TUG subtasks. A novel approach by You et al. [35] incorporated the TUG test to extract features from the whole-body movement (25 joints) and discovered that joints of the legs and hands contribute to classification.

Zhang et al. [36] employed single and dual gait tests and extracted gait cycles from each gait task, other features such as mean speed and standard deviation of speed, mean step length, mean gait cycle time, and mean acceleration were extracted from the whole gait task.

Kondragunta et al. [40] applied 3D pose estimation to obtain the subject's pose and extracted ankle coordinates. They calculated the 3D distance between the two feet for each frame using Euclidean measurement and applied Dynamic Time Warping to identify important temporal parameters.

Aoki et al. [41] demonstrated that dual-task features obtained a better performance compared to single gait features for classification. The study utilized whole-body movement and by employing a time-frequency analysis on a sequential data set comprising 3D coordinates of bodily joints derived a comprehensive set of gait features.

Nielsen et al. [43] recorded the dual-task in seconds and the time taken to complete the TUGdt is fed into LR. The features considered in Ng et al. [44] encompass the TUG metric evaluating functional mobility, rapid gait speed (GS), knee extension strength (KES), and performance-based mobility assessment.

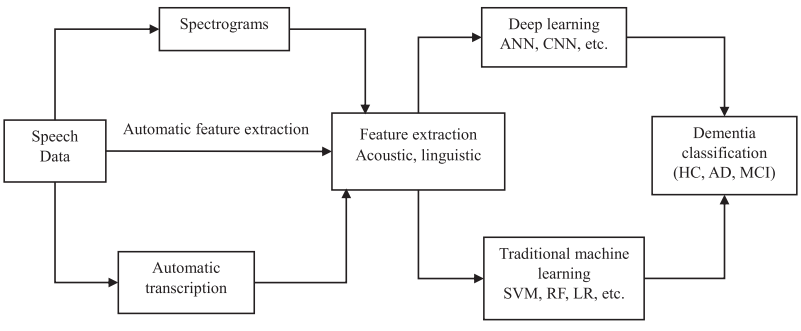

Fig. 5. Overview of identified machine learning methodology for speech analysis. HC, healthy controls; AD, Alzheimer's disease; MCI, mild cognitive impairment; SVM, support vector machine; LR, Logistic regression; RF, Randomforest; ANN, Artificial neural network; CNN, Convolutional neural network.

SPEECH ANALYSIS FOR DEMENTIA DETECTION USING MACHINE LEARNING

Speech analysis has the potential to offer a cost-effective, simple to implement, and accurate way for early onset dementia detection [29]. This section demonstrates a comprehensive evaluation of dementia detection from speech using machine and deep learning models. The overview of the machine learning methodology for speech analysis is shown in Fig. 5. Machine and deep learning algorithms will be discussed first followed by the features employed in the selected studies. All machine learning and deep learning models in the selected studies used supervised learning. The models used for deep learning are artificial neural networks (ANNs), multilayer perceptron (MLP), recurrent neural network (RNN), Bidirectional Encoder Representations from Transformers (BERT), A Lite BERT (ALBERT), and Robustly Optimized BERT Pretraining Approach (RoBERT). Moreover, methods used for machine learning are mostly SVM, RF, and LR. All algorithms distribution across studies can be seen in Fig. 6.

Deep learning

Deep learning consists of multiple layers of artificial neurons to learn intricate representations of the input data. Among the search results, deep learning algorithms are used for classification ($N = 14$). In Vetráb et al. [46] deep learning is used for feature extraction and classification is performed with SVM ($N = 14$).

Themistocleous et al. [47] trained 10 deep neural network architectures based on MLP and the best performance was achieved utilizing 5-fold cross-validation to distinguish between MCI and HC with an accuracy of 83%. The authors suggested that the acoustic features that encompass vowel duration, vowel formants, and fundamental frequency along with gender and age as predictors resulted in important information for the detection of MCI.

Al-Atroshi et al. [48] introduced an automatic speech recognition model utilizing the Gaussian mixture model and deep belief network GMM-DBN and SoftMax classifier. The Method depends on the Gaussian Mixture Model (GMM) for automatic speech recognition (ASR) of the speech. The Deep Belief Network (DBN) model is used to extract feature vectors from identified speech data. The SoftMax classifier was built for the classification of MCI and AD disorders in the used speech signals. GMM-DBN achieved an F1-score of 90.19% and an accuracy of 90.28% in the detection of multiple classes, and on the binary classification, it achieved an accuracy of 86.76%.

A CNN containing three convolution blocks and three fully connected layers was employed by Chau et al. [49]. Voice spectrograms are directly used as input to CNN for the classification of MCI. The study explored the optimal threshold for MCI classification and achieved a precision of 75% and recall of 82% when the threshold was optimum.

Bertini et al. [50] implemented an autoencoder trained on spectrograms of the speech data and used MLP for classification. The MLP consists of 4 hidden layers along with 128 hidden rectified linear units. The authors reported an accuracy of 90.57% in the binary classification and 85.39% in the multiple classification (HC, MCI, eD), and outperformed the techniques relying on manual transcription and annotation of speech.

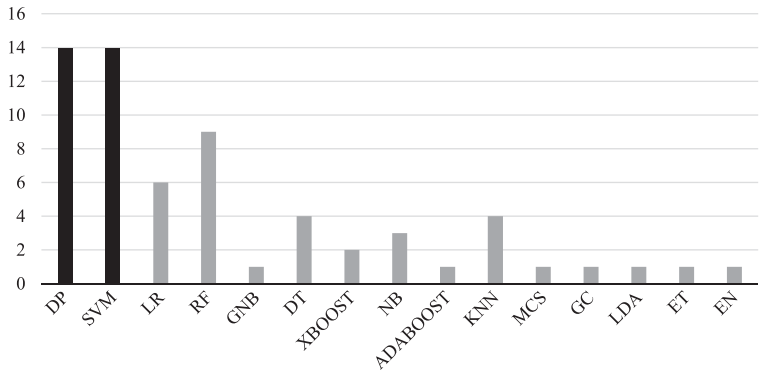

Fig. 6. Frequency of machine learning algorithms using speech analysis for dementia prediction. SVM, support vector machine; DP, Deep Learning; LR, Logistic regression; DT, Decision trees; RF, Random forest; NB, Naïve Bays; KNN, K-Nearest Neighbor; GNB, Gaussian Naïve Bayes; XBOOT, Extreme Gradient Boosting; ADABOOST, Adaptive Boosting; MCS, Monte Carlo Simulation; GC, Gaussian Classifier; LDA, Linear Discriminant Analysis; ET, Extra Trees; EN, Ensemble Model. The black color shows the most frequent algorithms appearance in the selected studies.

Sadeghian et al. [51] combined acoustic, MMSE test results, and linguistic features for the detection of AD from HC. A total of 236 (including four demographic) features were extracted with the aim of selecting the most informative one. According to the study, the MMSE score is the most informative parameter (accuracy of 71% alone), when combined with other features yielding the best performance. Genetic Algorithm-Support Vector Machine (GA-SVM), RF, and ANN specifically MLP were implemented, and the best accuracy was obtained with the neural network. The accuracies of automatic (94.4%) and manual transcription (95.8%) of NN were almost similar. However, Sadeghian et al. [52] demonstrated that acoustic and linguistic features were able to discriminate even if the MMSE score was not considered. The authors built MLP with one hidden layer composed of 25 neurons and the sigmoid as an activation function was used. The MLP attained an accuracy of 94.4% with five features (MMSE score, race, fraction of pauses greater than 10 sec, fraction of speech length that was pause, and the Linguistic Word Inquiry Count (LIWC)). This study shows that manual and automatic transcription produce similar accuracies of 91.7% and 93.1% respectively.

Sangchocanonta et al. [53] implemented MLP to classify between MCI, AD, and HC. The authors developed two new picture description tasks and transcribed the speech automatically. Part-of-Speech (PoS) tagger was used to create 47 types of words and grouped them into seven parts of speech. Among DT, RF, NB, LR, and Adaboost, MLP obtained the best results with an accuracy of 86.67%, sensitivity of 90.00%, and specificity of 80.00%.

Chien et al. [54] used a variant of RNN (biGRU) for the automatic assessment of AD from HC by employing a feature sequence generator. The study obtained an AUC value of 83.8% after fine-tuning the model. The authors conducted both manual and automatic transcription and the AUC was 80.0% for automatic transcription and 83.8% for manual transcription.

Traditional machine learning methods

Traditional machine learning approaches for dementia detection have been implemented after feature engineering of the speech signal. Several classifiers have been utilized to achieve a better performance in dementia disease classification. The most frequent model present in the selected studies was SVM ($n = 14$), see Table 4. which shows the occurrence of machine learning models in the selected studies. This section aims to outline the machine learning techniques and the most significant features used.

Table 4
Number of Machine learning methods used for dementia detection through speech analysis

Model	Number of studies
Support Vector Machines	14
Random Forest	9
Linear Regression	6
K-Nearest Neighbors	4
Decision Trees	4
Naïve Base	3
XBoost	2

Application of support vector machines

An SVM was implemented by López-de-Ipiña et al. [55] for the detection of dementia. The SVM showed superior performance over CNN, KNN, and MLP with a classification error rate (CER) of 5% using the best 25 features. The second-best algorithm was CNN with a CER of 8%. Classical Features (spectral and time domain), Perceptual Features (MFCC), and nonlinear features (fractal and entropy features) were selected through a hybrid approach by nonparametric Mann–Whitney U test and SVM evaluation. Feature selection and normalization are done with the WEKA tool.

Vetráb et al. [46] aimed to classify MCI from HC by implementing sequence-to-sequence deep autoencoders for feature extraction and linear SVM for classification. The best accuracy achieved in this study was 72% and the corresponding AUC value was 0.763, which was obtained by clipping the power levels below a certain dB threshold (-75 dB). The study found that the sequence-to-sequence autoencoders outperformed x-vectors, even though they were trained on more data.

Gosztolya et al. [56] developed an SVM with a linear kernel for identifying MCI and AD based on spontaneous speech using acoustic and linguistic features. The authors evaluated the usefulness of both the acoustic and linguistic features. Applying only demographic features the accuracy obtained was 40%. Using only acoustic features, they were able to separate different groups with an accuracy of 74-82% (for the 3-class HC, MCI, mAD) and they got similar accuracy results by using only linguistic features. By combining both acoustic and linguistic features the performance improved with ranges of

80% -86%, and 78-86% of the accuracy and recall respectively.

Yu et al. [57] implemented an SVM model to monitor cognitive impairment by analyzing speech from remotely collected cognitive tests. An extended set of features including articulation, phoneme characteristics, and pitch were extracted. The SVM model achieved the highest performance (AUC = 77%) when the features were combined with the cognitive test score. However, using speech features alone, the AUC for distinguishing cognitive impairment from healthy individuals was 74%, while for the cognitive test, it was 45%. The best performance features were those combined with cognitive tests.

König et al. [58] employed an SVM for classification between HC, MCI, and AD. The study used the word count, clusters, and switches features extracted from the semantic verbal fluency speech task. The classification performance for AUC was HC versus AD = 93.9%, HC versus MCI = 75.8%, and AD versus MCI = 85.9%. The study performed both manual and automatic transcription. The authors aimed to examine whether automatic analysis of the semantic verbal fluency task can provide as accurate classification as the manual analysis and thus aid the cognitive impairment screening. For both the comparisons of HC versus AD and HC versus MCI, the models trained using automated features and manual features exhibited similar performances in terms of AUC, with values of 72.3% versus 75.8% and 95.3% versus 93.9%, respectively. However, in the context of MCI versus AD, the AUC of models trained on automated features showed a decline, with values of 85.9% compared to 77.4% for models utilizing manual features.

König et al. [59] implemented an SVM for differentiation between different classes of dementia (subjective cognitive impairment (SCI), MCI, AD, and Mixed). The study used different speech tasks, including verbal fluency, picture description tasks, free speech tasks, and counting down. Vocal markers are derived using speech signal processing. Then a subset of the most relevant vocal markers was selected by the Mann-Whitney test. The algorithm obtained the following accuracies: SCI versus AD = 92%, SCI versus MD = 92%, SCI versus MCI = 86%, MCI versus AD = 86%, and 82% for MCI versus Mixed cases. In scenarios involving multiple classes, the results stand at 78% accuracy for differentiating SCI, MCI, and AD, 74% for SCI, MCI, and Mixed; and 75% for SCI, MCI, and an alternate category (either AD or Mixed).

Tóth et al. [60] implemented an SVM to detect MCI from HC based on the analysis of spontaneous speech where acoustic parameters were extracted. The extracted acoustic parameters comprised of hesitation ratio, speech tempo, length, number of silent and filled pauses, and length of utterance. Those parameters showed a correlation to MCI, so they could be considered as biomarkers. The study demonstrated that a fully automated process was able to separate the classes with an F1-score of 78.8%. This study performed both automatic and manual transcription. SVM and RF achieved an AUC of 63.9% for automatic and 73.4% for manual and 67.6% for automatic and 73.4% for manual, respectively.

Metarugcheep et al. [61] incorporated an SVM to distinguish between MCI and HC. Feature extraction based on phonemic clustering and switching through the phonemic verbal fluency (PVF) task including acoustic, semantic, and grouping of words. In this study, Silence-based features, Similarity-based features, and cluster features were the three categories of features used for analysis. Feature selection was done by means of the chi-square test which helps in highlighting the most important features. The study used three classifiers (extreme gradient boosting (XGBoost), SVM, RF) and in terms of performance, SVM performed the best with an AUC of 73.3%.

Application of logistic regression

Shimoda et al. [62] developed machine learning algorithms for the prediction of AD from HC based on vocal features extracted from daily conversation. The proposed machine learning algorithms include XGBoost, RF, and LR in which LR performed the best in terms of AUC = 89.3%. When evaluating each participant as one data point and calculating the mean predictive value of their audio files, the AUC values for XGBoost, RF, LR, and the cognitive test questionnaire were found to be 1.00. implying perfect predictive performance.

Nagumo et al. [63] investigated whether acoustic features could discriminate HC from MCI and global cognitive impairment (GCI). The features included temporal attributes encompassing duration of utterance, number, length of pauses, and spectral features represented by fundamental frequency (F0), first formant (F1), and second formant (F2). LR was employed and achieved an AUC of 61% for the classification of HC versus MCI, AUC of 67% for HC versus GCI, and 77% for MCI with GCI.

Munthuli et al. [64] devised a new speech task called Thai Language and Cognition Assessment (TLCA) that is specified for the Thai language and culture for screening of MCI and AD from HC. The TLCA contains 21 tasks out of which 16 are speech tasks. 988 features were extracted with the OpenSMILE [65]. Feature selection using correlation-based and best-first search methods to choose the most significant features was performed. LR demonstrated superior performance compared to MLP and RF, achieving accuracy of 100%, and 100% sensitivity in the multiclass classification (HC, MCI, AD) and accuracy of 93.33% and 100% sensitivity in the binary class (HC, non-HC).

Application of K-Nearest Neighbor (KNN)

KNN is a widely used machine learning algorithm for both classification and regression. The principle of the KNN algorithm is that it classifies points based on the class of it is nearest neighbors. The distance is calculated between the data points using commonly used distance measures such as Euclidean or Manhattan distance [66].

Mirzaei et al. [67] introduced a two-phase approach for feature selection for the detection of AD using voice features to optimize the prediction accuracy. Temporal and acoustical features are used for classification. The features were normalized based on the z-score. Moreover, Mel-Frequency Cepstral Coefficients (MFCCs) and Filter Banks Energies (FBEs), were evaluated. The highest accuracy was achieved with a KNN of 62% after feature selection.

Nasrolahzadeh et al. [68] used higher-order spectral features for classifying AD subjects compared to HC. Information gain criteria were adopted for feature selection. KNN, SVM DT, and NB were implemented and achieved comparable results in terms of accuracy with 97.71, 96.46, 94.18, and 95.94%, respectively.

Application of ensemble model

Ensemble models are methods that combine the prediction of multiple algorithms. This method allows us to get the strength of the combined models to produce more accurate and robust results. The main idea behind ensemble models is that multiple weak learners merge to form a strong one as a result it significantly improves the performance [69].

The parallel classification model consisting of KNN and SVM was adopted by You et al. [70] to discriminate between high-risk and low-risk dementia. The study utilized 282 paralinguistic features for analysis and a final set of 27 features including MFCC coefficients, pitch, energy, skewness, spectral centroid, and memory test result, yielded the best classification accuracy. The Mann-Whitney U-test, Neighborhood Component Analysis (NCA), and ReliefF were used for feature selection. The ensemble classification result of the SVM and KNN was 94.7% with only paralinguistic features and 97.2% using both paralinguistic and episodic memory features.

Other traditional machine learning classifier. Weiner et al. [71] used both linguistic and acoustic features extracted from bibliographic interviews and speech recordings to predict the development of dementia. The Gaussian classifier was implemented to accomplish unweighted average recall (UAR) of 73.3% and 75.7% for the prediction of dementia five and 12 years ahead.

De Looze et al. [72] explored the relationship between speech chunking and cognitive impairment in individuals with MCI and AD. Temporal features were extracted from the reading task, including speech chunks, pauses, dysfluencies, rates, and duration. The authors found mild-to-moderate AD is related to slower speech with more chunks, pauses, and disfluencies. Linear discriminant analysis (LDA) was employed to discriminate between groups and achieved an AUC of 75% to discriminate between healthy and MCI individuals, an AUC of 90% between MCI and AD, and between AD and HC, with an AUC of 94%.

Soroski et al. [73] compared automatically transcribed speech using Google speech-to-text software to manually corrected transcripts to see the effectiveness of automatic transcription for dementia classification. The automatic transcription showed higher confidence scores and lower error rates for speech data from HCs compared with patients. Machine learning classification accuracies with the manually corrected transcripts outperformed the automatically generated transcripts for the picture description and recalling a pleasant experience task, but there is no difference for the reading task. The study found that adding manual pauses to transcripts had no impact on classification performance.

Liang et al. [74] explained that the machine learning classification algorithms built using the extracted features achieved an accuracy of 68% using features fusion and 80%-90% using DT and RF models with selected features. The Root-Mean-square Error (RMSE) result of the regression model was 3.53.

Table 5
Summary of the selected speech studies

Author	Language	Task	Dataset	Demographic	Features	Validation method	Algorithm	Best Accuracy
Soroski et al. [73] (2022)	English	PD, RT, Recall a pleasant life experience	$N = 149$ (AD = 44, MCI = 20, SMC = 8, HC = 77)	mean age (71.9 ± 8.9), age range (53–96) Female=89, Male = 60	Linguistic features, Acoustic features	stratified 10-fold cross-validation	a) LR, b) RF, c) GNB, d) BERT	a) AUROC = 74.3%
Liang et al. [74] (2021)	English	30 pre-selected voice commands	$N = 40$ (HC = 18, MCI = 22)	Male (HC = 9, MCI = 7), female (HC = 9, MCI = 17) Age range (65–80)	163 unique command-relevant features were extracted	Leave-One-Subject-Out-cross-validation	a)1NN/2NN, b) SVM, c) DT and RF	c) Acc = 80%–90%.
Chau et al. [49] (2022)	Mandarin or Taiwanese	Conversation during MSE	$N = 120$ (HC = 84, MCI=36)	Male=46 (MCI = 17, HC = 29), Female = 74 (MCI = 19, HC = 55) Age range (61–90)	Voice spectrogram images	stratified split	CNN	precision=75% Recall=82%
Al-Atroshi et al. [48] (2022)	Hungarian	SS	$N = 75$ (HC = 25, MCI = 25, AD = 25)	The class labels are matched based on gender, age, and education.	Feature vectors	5-fold cross-validation	GMM-DBN for ASR and feature extraction, SoftMax classifier	Acc = 90.28%
Vetráb et al. [46] (2022)	Hungarian	SS	$N = 50$ (HC = 25, MCI = 25)	The groups are matched in terms of age, gender, and education	Compressed hidden states, Number of chunks, I-vectors, X-vectors	25-fold stratified cross-validation	Sequence-to-sequence deep autoencoder for feature extraction, linear SVM	Acc = 72%
Bertini et al. [50] (2021)	Italian	SS	$N = 96$ (HC = 48, MCI = 32, eD = 16)	Male = 48, Female = 48, age range (50 to 75) Mean age (64.10 education>=junior high school	Log mel Spectogram	Train-test split method (80–20)	Autoencoder which a type of RNN (auDeep) and multilayer perceptron for classification	Acc = 90.57%

(*Continued*)

Table 5
(*Continued*)

Author	Language	Task	Dataset	Demographic	Features	Validation method	Algorithm	Best Accuracy
Shimoda et al. [62] (2021)	Japanese	Telephone of conversations of daily weekly for 1–2 months	$N = 123$ (HC = 99, AD = 24)	Mean age (75.65) Female=66, Male = 57 Age>=65	Vocal features	training-testing split (80/20	a) XGBoost, b) RF, c) LR	c) AUC = 89.3%
Sadeghian et al. [51] (2021)	English	PD	$N = 72$ (HC = 46, AD = 26)	Mean Age (74.04) Mean Years of education (13.48)	Acoustic and linguistic features, MMSE Result	10-fold cross-validation, LOO	a) RF-K-fold, b) RF-LOO, c) ANN-manual (MLP), d) ANN automatic (MLP), e) GA-SVM)	c) Acc = 95.8% d) Acc = 94.4%
Sangchocanonta et al. [53] (2021)	Thai	2 PDs (Thais-at-Home and Thai Temple Fair)	$N = 90$ (HC = 30, MCI = 30, AD = 30)	Male=28, female = 62 Age Range (57–86) mean age = 70.3	(PoS) tagging	5-fold cross-validation	a) DT, b) RF, c) NB, d) LR, e) MLP, f) Adaboost	e) Acc = 86.67%
Munthuli et al. [64] (2021)	Thai	TLCA (Memory, Language, cognition)	$N = 90$ (HC = 30, MCI = 30, AD = 30)	Male=33, Female = 57 Mean age (69.61)	Emobase feature set from OpenSmile that includes 988 features	5-fold cross-validation	a) MLP with fold 4, b) RF with fold 2, c) LR with 1-fold	c) Acc = 100.00%
López-de-Ipiña et al. [55] (2018)	Spanish	CVF, AN	$N = 100$ (HC = 62, MCI = 38)	Male=43, Female = 62 Mean age (n (56.94) Age range (39–79)	Linear and non-linear features	k-fold cross-validation with $k = 10$	a) CNN, b) SVM, c) KNN, d) MLP Feature selection (MWU test)	b) CER = 5%
Nagumo et al. [63] (2020)	Japanese	RT	$N = 8{,}779$ (HC = 6,343, MCI = 1,601, GCI = 367, MCI w/ GCI = 468	Male=3,907, Female = 4,872 (mean age 74.02) Age range (65–96)	Acoustic features and spectral features	Nested cross-validation outer (3-fold cross-validation) inner (5-fold cross-validation)	LR	AUC-ROC = 77%
Weiner et al. [71] (2019)	Germany	biographic interviews	t+5y (HC = 57, AACD = 14, AD = 6) t+12y (HC = 41, AACD = 10, AD = 6)	–	Linguistic and acoustic features	leave-one-person-out cross-validation	MCS	UAR (75.5%) For predicting development of dementia 12 years later

Gosztolya et al. [56] (2018)	Hungarian	SS	$N = 75$ (HC = 25, MCI = 25, mAD = 25)	All groups were matched for age, gender and education. Mean age (72.36)	Acoustic and linguistic features, combined (acoustic+linguistic)	5-fold nested Cross-validation	SVM with linear kernel using SMO implementation in the WKA machine learning library	Combined features (Acoustic + linguistic:)) = 80% – 86%
Mirzaei et al. [67] (2018)	French	RT	$N = 48$ (HC = 16, MCI = 16, AD = 16)	Mean age (76.06) Age=>60	temporal and acoustical voice features	8-fold cross-validation	a) KNN, b) SVM, c) DT	a) Acc = 62%
Nasrolahzadeh et al. [68] (2018)	Persian	SS	$N = 60$ (HC = 30, AD = 30)	Mean age (74.22) Age Range (56–98) Male=31, Female = 29	higher order spectral features (bispectrum)	Ten-fold cross validation	a) KNN, b) SVM, c) NB, d) DT	a) Acc = 97.71%.
Themistocleous et al. [47] (2018)	Swedish	RT	$N = 60$ (HC = 30, MCI = 25)	Mean age (69) Age Range (55–79) Male=32, Female = 23	Acoustic features, Sociophonetic Features (age, gender)	5-fold cross-validation, 90/10 split	DNN (10 NNs with different number of hidden layers based on MLP)	Acc = 83%
Yu et al. [57] (2018)	English	Animal fluency task, East Boston memory test	$N = 282$ (HC = 260, Amnestic MCI/SD=15, Amnestic MCI/MD=5, Dementia = 2)	Age>75 (diagram)	Phoneme-based, pitch, articulation, Pseudo-syllable-based, Formant-based, Cognitive test score	LPOCV	a) SVM, b) GC	a) AUC = 77%
König et al. [58] (2018)	French	SVF	$N = 95$ (HC = 24, MCI = 47, ADRD = 24)	Mean age = 76.80 Male=36. Female = 59 Edu (HC = 10.50, MCI = 10.81, ADRD = 9.75)	Word count, clusters, Mean Cluster Size, number of switches	LOOCV	a) SVM, b) normative equations	a) AUC (HC versus AD = 93.9%, HC versus MCI = 75.8%) AD versus MCI = 85.9%

(*Continued*)

Table 5
(*Continued*)

Author	Language	Task	Dataset	Demographic	Features	Validation method	Algorithm	Best Accuracy
König et al. [59] (2018)	French	Verbal fluency, PD, Free speech task, counting down	*N* = 165 (AD = 27, MCI = 44, Mixed = 38, SCI = 56)	Mean age = 76.2 Age<=65 Male=57. Female = 108	Vocal markers	Repeated subsampling	SVM	Acc = 92%
Tóth et al. [60] (2017)	Hungarian	SS (recalling two films)	*N* = 86 (HC = 38, MCI = 48)	Mean age = 68.60 Range (55–93) Male=29. Female = 55	Acoustic features	LOOCV	a) NB, b) RF, c) SVM with WEKA tool	b) AUC = 73.4% c) AUC = 73.4%
López-de-Ipiña et al. [79] (2017)	multicultural and multilingual database	a) CVF, b) PD, c) SS	a) PGA-OREKA (HC = 62, MCI = 38) MINI-b) PGA (HC = 12, AD = 6) c) AZTIAHORE (HC = 20, AD = 20)	a) PGA-OREKA (Male = 53, Female = 57) MINI-b) PGA (Male = 5, Female = 13) c) AZTIAHORE (Male = 19, Female = 21) Age range (39–88)	Linear and non-linear features	10-fold cross-validation	a) MLP, b) CNN	b) Acc = 95%
Sadeghian et al. [52] (2017)	English	PD	*N* = 72 (HC = 46, MCI = 26)	Mean age = 74.04 Edu mean (13.48)	Acoustic and linguistic features, MMSE test	10-fold cross-validation	MLP	Acc = 94.4%
Kálmán et al. [75] (2022)	Hungarian=33, English = 33	SS	*N* = 66 (HC = 39, MCI = 27)	Male=18. Female = 48 Mean age (72.62) Edu mean (14.88) Age>=60	Temporal parameters			
De Looze et al. [72] (2018)	English	RT	*N* = 70 (HC = 36, MCI = 16, AD = 18)	Mean age (72.44) Edu mean (13.30) Male=36. Female = 34	temporal features	LOOCV	LDA	ROC AUC = 94%
Chien et al. [54] (2019)	Taiwanese	Fluency test, PD, logical memory test	*N* = 80 (HC = 40, AD = 40)		Feature sequence generator	Cross validation	RNN	AUC-ROC 83.8%

Schäfer et al. [76] (2022)	a) Swedish = 404, b) Dutch = 121	SVF, RAVLT	a) N = 121 (SCI = 69, MCI = 52), b) N = 404 (CI = 356, MCI = 48)	a) Mean age = 62.20 b) Mean age = 75	ki:e SB-C composed of more than 50 features derived from the cognitive tests and only 27 features were selected	LOOCV, grid search for hyper parameter tuning	a) SVM, b) RF	a) AUC = 81%
Reeves et al. [77] (2022)	English	ND (video clip was shown to the subject)	N = 52 (HC = 16, SCD = 14, MCI = 13, Dementia = 9)	Mean age = 82.5 Range (65–99) Male=22. Female = 30	Linguistic features	–	Backward stepwise logistic regression	AUC = 89%
Li et al. [78] (2020)	English	Daily activity, room environment, PD	B-SHARP Dataset N = 326 (HC = 185, MCI = 141)		Linguistic features		a) MLP, b) RoBERTa, c) BERT, d) ALBERT, e) EN	e) 74.1%
Metarugcheep et al. [61] (2022)	Thai	PVF	N = 100 (HC = 59, MCI = 41)	Male=17. Female = 83	Acoustic features, linguistic features	10-fold cross-validation	a) SVM, b) XGBoost, c) RF	a) Acc = 73.3%
You et al. [70] (2019)	English	episodic verbal memory test (LOGOS)	N = 4,423 (high versus low dementia risk)	Age Range (55–77)	paralinguistic features, episodic memory test features	10-fold cross validation	a) SVM, b) KNN, c) EN of KNN and SVM	c) Acc = 97.2%

PD, Picture Description Task; RT, Reading Task; N, number; AD, Alzheimer's disease; MCI, mild cognitive impairment; HC, healthy control; Acc, accuracy; SMC, subjective memory complaints; LR, Logistic Regression; RF, Random Forest; BERT, Bidirectional Encoder Representations from Transformers SS, spontaneous speech; GNB, Gaussian Naive Bayes; 1NN, One Hidden-Layer Neural Network; 2NN, Two Hidden-Layer Neural Network; DT, Decision Tree; SM, Soft Max; eD, Early Dementia; CNN, Convolutional Neural Network; SVM, support vector machine; XGBoost, Extreme Gradient Boosting; MLP, Multilayer Perceptron; ANN, Artificial Neural Network: NB, Naïve Bays; CVF, Categorical Verbal Fluency Task; AN, Animal Naming; MWU, Mann–Whitney U test; CER, Classification Error Rate; AUC-ROC, Area Under the Receiver Operating Characteristic Curve; AACD, Aging-Associated Cognitive Decline; UAR, Unweighted Average Recall; MCS, Monte Carlo Simulation; LPOCV, Leave-Pair-Out Cross-Validation; MCI/SD, Single Domain; MCI/MD, Multiple Domain; CG, Gaussian Classifier; ADRD, Alzheimer's disease and related disorders; LOOCV, Leave-One-Out Cross Validation; SVF, Semantic Verbal Fluency; CVF, Categorical Fluency Task; RAVLT, Rey Auditory Verbal Learning; ND, Narrative Description; PVF, Phonemic Verbal Fluency; mAD, early stage AD.

The authors also analyzed the contribution of each feature set to the model output, revealing the most impactful features for cognitive assessments. Overall, the study demonstrates the potential of using Voice-Assistant Systems (VAS) for home-based cognitive assessments and early detection of cognitive decline in older adults living alone.

Study characteristics

The most used classification classes in the literature were between HC and MCI as seen in Table 6. In terms of languages used Table 7. shows the languages used in the included studies in which English language was the most used. Regarding the study design, 24 are cross-sectional studies, five are longitudinal studies, and two are retrospective studies.

Table 6
Summary of classification classes results of the machine learning methods

Type of Study	Number of studies
MCI	11
AD	5
MCI and AD	6
AD, MCI, SMC	1
MCI and early dementia	2
MCI, GCI, GCI with MCI	1
Amnestic MCI dementia single and multiple-domains	1
AADC and AD	1
MCI, AD, mixed and SCI without healthy control	1
SCI, CI MCI	1

AD, Alzheimer's disease; MCI, mild cognitive impairment; HC, healthy control; SCI, subjective cognitive impairment; SMC, subjective memory complaints; CI, cognitive impairment; AACD, aging-associated cognitive decline; GCI, Global Cognitive Impairment.

Table 7
Distribution of languages in the included speech studies for dementia prediction

Language	Number of studies
English	9
Hungarian	4
French	3
Thai	3
Japanese	2
Italian, Spanish, German, Persian, Swedish, (Mandarin or Taiwanese), Taiwanese	1 study in each language
Multicultural and multilingual database	1
Swedish or English	2
Swedish, Dutch), (Hungarian, English)	2 studies used two databases

Feature characteristics

Features are crucial for obtaining optimal classifier performance to discriminate between different levels of impairment. Many features have been evaluated in the literature including linguistic features, acoustic features, spectral features, or a combination of both linguistic and acoustic features for dementia detection from speech. For selecting the best features, several studies used the nonparametric Mann–Whitney U test for feature selection [55, 59, 70, 79].

In the literature, feature vectors were extracted from spontaneous speech tasks. Al-Atroshi et al. [48] used The DBN model to extract the feature vector from the spontaneous speech task. Nasrolahzadeh et al. [68] used the bispectral structure of spontaneous speech. Moreover, to measure the performance of the features, they used the gain ratio method. Vetráb et al. [46] employed sequence-to-sequence deep autoencoder to extract compressed hidden states, number of chunks, I-vectors, and X-vectors from spontaneous speech tasks.

Schäfer et al. [76] extracted over 50 features from two neuropsychological assessments, Rey- Auditory Verbal Learning Test (RAVLT) and Semantic Verbal Fluency, and 27 features were selected for analysis. The study grouped the extracted features into three neurocognitive subdomains, encompassing learning and memory, executive function, and processing speed, and calculated a single global composite score from these subdomains to represent an overall measure of cognitive function based on speech patterns.

Acoustic features

Acoustic features describe the characteristics of the speech sound. Acoustic features include pitch, intensity, duration, MFFCC, prosodic features, and formants. Acoustic features are the most popular in search results.

Liang et al. [74] used a voice assistant system and extracted 163 features from each participant performing 30 per selected voice command. Two fusion strategies, early fusion, and late fusion are employed. In early fusion, all feature values are combined into a single vector from the outset. In contrast, the late fusion strategy involves training separate classifiers and then combining their outputs. Mirzaei et al. [67] Implemented a two-stage feature selection method. The study extracted features,

including prosodic, temporal, quality-related measures, MFCCs, and FBEs, were extracted from speech samples. The wrapper feature selection method was employed to select the most relevant features for disease prediction.

De Looze et al. [72] employed temporal features, including sentence duration, speech chunk characteristics, pause duration, articulation rate, speech rate, and dysfluencies, were automatically extracted from annotations, and a mixed model approach with backward elimination was employed for feature selection in analyzing the effects of cognitive status and linguistic demand on speech. Nagumo et al. [63] extracted acoustic features, mainly they focus on the temporal aspect of the speech. Kálmán et al. [75] used automatic speech recognition to assess fifteen temporal parameters, including speech tempo, articulation tempo, silent pause duration rate, total pause duration rate, silent pause average duration, and total pause average duration, revealing significant differences between individuals with MCI and HCs in both English and Hungarian-speaking samples.

Shimoda et al. [62] utilized the open software tool PRAAT [80] to extract 60 vocal variables, including mean, median, minimum, maximum, 0.15 percentile, and 0.85 percentile, and successfully achieved accurate prediction of AD using these vocal features extracted from daily conversations.

Munthuli et al. [64] utilized openSMILE's Emobase feature set, comprising 988 features derived from 19 functionals of 26 low-level descriptors extracted from 1,440 speech signal files, encompassing attributes such as intensity, loudness, MFCCs, and Zero Crossing Rate. König et al. [59] incorporated vocal features extracted from speech recordings, including low-level markers such as silence and voice duration, and selected relevant features using the Mann-Whitney test based on their significance in discriminating between cognitive different impairment groups (SCI, MCI, AD, Mixed)

Themistocleous et al. [47] used both acoustic and Sociophonetic (age and gender) Features. The features included vowel formants (F1 to F5) measured at multiple positions, fundamental frequency (F0) characteristics (mean, minimum, and maximum), vowel duration, participants' gender, and age, resulting in 24 acoustic and sociophonetic features.

Tóth et al. [60] extracted acoustic features including speech rate, articulation rate, number of pauses, total length of pauses, and pause rate. Duration of the utterance, and average length of pauses. The study showed that those features are significantly different between MCI and HC.

Linguistic features

Linguistic features describe the structure of the spoken language. It includes word count, syntax, fluency, sentence complexity, speech fillers, vocabulary diversity, and hesitation.

Part of speech tagger (POS) is performed by Sangchocanonta et al. [53] to group words into seven different categories. POS tool is used in natural language processing to group words into their grammatical categories (e.g., verb, noun, etc.) [81]. König et al. [58] utilized features of word count, semantic clusters, mean cluster size, and the number of switches. Reeves et al. [77] analyzed linguistic features including content efficiency, speech rate, part of speech usage, and unique word counts. The study found that linguistic features varied between subjects with MCI and SCI, and thus can be used for the detection of cognitive impairment. Moreover, linguistics showed a slight improvement in the sensitivity and specificity of the narrative description test score. Using both the narrative description test and linguistic features were able to predict cognitive impairment.

Chien et al. [54] used feature sequence generator. The feature representation in the study consisted of a sequence of tokens, primarily using syllables of Mandarin Chinese, which capture various speech characteristics including pausing, repeating, and disfluency, and are generated using a Convolutional Recurrent Neural Network (CRNN) trained on speech data from multiple datasets.

Hybrid features

Several studies used different combinations of features for dementia prediction. Studies showed that the best performance is by a combination of cognitive test scores with other types of features [57, 77].

Yu et al. [57] utilized features including phonemic, pseudo-syllable, and articulatory measures. The study found that combining cognitive test scores with speech parameters, particularly the animal fluency score, improved cognitive impairment detection compared to using either one in isolation. This implies that incorporating both cognitive test scores and speech parameters may provide complementary information for more accurate cognitive impairment detection.

Sadeghian et al. [51] found that the most informative single parameter among the 237 potential features, including four demographic factors and MMSE scores, is the MMSE score itself, with acoustic and linguistic features such as pause counts, pause durations, pitch variations, vocabulary richness, word

type usage, sentence complexity, content accuracy, and idea density also being considered.

Weiner et al. [71] employed a nested forward feature selection approach to effectively trim down the dimensionality of a substantial 444-dimensional feature vector. This feature reduction process was carried out for the purpose of dementia prediction, utilizing a combination of linguistic and acoustic features.

Sadeghian et al. [52]. utilized a combination of acoustic features from speech, linguistic features extracted from transcripts, and MMSE scores achieving the best results when considering all features, including MMSE scores, with a detection accuracy of approximately 94%.

Metarugcheep et al. [61] utilized a combination of silence-based, similarity-based, and cluster features extracted from audio recordings and transcribed files, with feature selection methods based on their relevance to distinguish between individuals with MCI and HC. Feature selection was done using a chi-square test, based on their relevance.

Gosztolya et al. [56] experiment on different combinations of features. The study attained accuracy of 74% to 82% using only acoustic features, similar scores with linguistic features, and improved accuracy of 80% to 86% when combining both feature types, with corresponding F1 values between 78% and 86%.

Soroski et al. [73] used both linguistic and acoustic features extracted from the picture description, reading, and experience description tasks. They utilized the correlation feature selection method to identify the most relevant features.

Speech analysis approach

Various studies perform different methods for analyzing the recorded speeches. In the literature review, the studies that involve manual transcription are excluded. It is seen from Table 8 that several studies relied on automatic transcription, while others combined both automatic and manual and compared their result. Additionally, a few studies utilized spectrograms, and others adopted automatic extraction of features.

Spectrograms are visual representation of the audio signal. Some studies have utilized them for the classification of different types of dementia as shown in Table 8. Spectrograms provide information about the frequency and time domain characteristics of the speech data, enabling the extraction of features that identify the subtle changes in speech.

Chau et al. [49] adapted discrete Short-Time Fourier Transformation (STFT) for building the spectrogram. The audio features were extracted from each 4-second sample using STFT, which involves applying a window function to the signal over short time intervals and then performing a Fourier transform on each of these windows to obtain the spectrogram representation of the audio data. Nasrolahzadeh et al. [68] employed the Fast Fourier Transform (FFT) for bispectrum estimation and bispectral features were extracted. Vetráb et al. [46] applied 128 log-scale Mel-spectrogram filters. The window width for the spectrogram was 0.08 ms, and there was a 0.04 ms overlap.

Automatic transcription plays an important role in the field of dementia detection from speech. This technology makes use of advanced machine learning algorithms and automatic speech recognition systems to convert speech into text with remarkable accuracy. In their research [52, 75], they developed their own ASR.

Studies utilized Google speech-to-text services for transcription [53, 59, 73]. In Liang et al. [74] study the transcription process involves the use of both Alexa's ASR system and the open-source Wav2vec ASR to generate transcripts from Alexa audio recordings. Automatic transcription was performed with the help of Kaldi software [51, 52]. Schäfer et al. [76] automatically process speech using a proprietary speech analysis pipeline. This pipeline includes automatic speech recognition for transcription and feature extraction, allowing them to analyze the verbal output of participants. Reeves et al. [77] used MacSpeech

Table 8
Transcription methods for the speech data. It shows studies used spectrograms, automatic transcription, both manual and automatic transactions, and only feature extraction.

Study	Spectrograms	Manual and Automatic	Automatic	Feature extraction
[51, 52, 54, 56, 58, 60, 71, 73]		✓		
[47, 48, 53, 74, 75, 76, 77, 78]			✓	
[46, 49, 50, 68]	✓			
[57, 62, 64, 55, 59, 61, 63, 67, 72, 79]				✓

scribe pro software and at least two reviewers edited for accuracy.

Several studies employed direct feature extracting to analyze the speech data. Shimoda et al. [62] and König et al. [59] employed vocal feature extraction using PAART software. A custom Python program for feature extraction was developed by Nagumo et al. [63].

DISCUSSION

"Early detection of dementia is crucial for improving the quality of life of old people, because it is a prerequisite for effective symptom-relieving interventions" [82]. This systematic literature review provides an in-depth summary of the existing body of knowledge for non-invasive, and easy-to-use methods for dementia detection using machine learning through gait and speech analysis.

In the comprehensive analysis of the literature pertaining to dementia-detection through gait analysis using machine learning, it is evident that limited research has been conducted. Specifically, there is a notable gap in studies utilizing gait analysis for this purpose, by employing convenient, less intrusive methods that do not cause discomfort to the subjects or distract their attention (e.g., a wearable sensor). Only nine studies were identified addressing this avenue of investigation. Thus, more research needs to be carried out in dementia detection through gait analysis, by employing techniques that do not interfere with or have contact with the subject.

Regarding machine learning models, few studies employed deep learning [34, 35, 36], and one reason could be that artificial intelligence algorithms require more data that is diverse and representative to learn patterns and enable the model to generalize well. The small amount of clinical data is due to privacy concerns or sometimes due to the lack of subjects concerning the same type of symptoms or disease. Neural Networks displayed better performance as compared to other traditional machine learning algorithms in the reviewed studies, hence deep learning models could be explored further to produce better classification results.

Analysis of dual-task and single-task for dementia detection provides valuable indications about the cognitive status of the subjects. However, dual-task showed better performance than single-task for dementia detection [41]. For gait analysis, more research is needed with a larger sample size that includes diverse classes of dementia and the use of new techniques to extract data. Zhang et al. [36] suggested combining gait analysis with other modalities such as speech and handwriting. By integrating speech and handwriting with gait, this approach could yield more accurate and reliable systems for dementia detection. Moreover, feature extraction from whole body movement, by including not only the lower body but also upper body movement could contribute to better classification performance [39, 35].

In the context of data-collection methods used in the literature for gait analysis, many studies adopted a Kinect camera which is a depth sensing technology and offers a simple and convenient method for collecting the data. However, gait analysis requires high precision and accuracy to detect subtle changes in gait movement. The Kinect sensor had limitations in the aspect of capturing fine details of human movement. Another limitation in the literature is that only one longitudinal study is implemented by Åhman et al. [8] as shown in Table 2.

This review for dementia detection from speech includes an extensive range of studies, that prove its promise to be used in clinical practice. The prevalence of speech studies is due to the ease of non-invasive data collection and significant advancements in natural language processing and transcription technology. Only studies with automatic transcription were considered because the field of machine learning, especially in the context of speech analysis, is moving towards end-to-end systems that can directly handle and process raw data. However, there are challenges to be discussed in this section. In terms of transcription methods, automatic transcription achieved comparable results to manual transcription [51, 52]. Moreover, Soroski et al. [73] recommended a human verification step with automatic transcription to ensure the reliability of the transcription. Hence, automatic transcription can offer promising benefits, as it requires less time and eliminates the tedious manual work. This finding is crucial as it suggests an automation process for speech recognition that can be utilized for the prediction of dementia.

Combining linguistic features such as part of speech, speech rate, and word frequency alongside acoustic features such as pitch, MFCCs, formants, and cognitive test scores tend to produce more promising results. Studies leveraging both linguistic and acoustic features [51, 52, 56, 57] tend to yield better results. This underscores the significance of adopting a comprehensive approach to speech anal-

ysis that considers not only the content and acoustic characteristics of speech but also the individual's cognitive test scores. By harnessing the richness of the combination of those features, researchers can develop more accurate and robust models for dementia detection.

Some studies suggested the use of speech result evaluation as a pre-screening tool and not to replace the clinical assessment [49, 53, 62]. Hence, interdisciplinary collaboration between machine learning experts, clinicians, and dementia researchers is crucial to ensure the development of clinically relevant and actionable models. Such collaborations can facilitate the development of more robust and user-friendly diagnostic tools that seamlessly integrate into existing healthcare systems. Health experts in the field of dementia can provide valuable insights into the clinical requirements, standards, and best practices for dementia diagnosis. Moreover, they can contribute to the validation and refinement of speech-based models by offering their expertise in subject assessment and monitoring.

It is worth noting that most of the examined studies predominantly focused on single-language datasets, and only one study adopted a multicultural and multilingual [79] database for dementia detection from speech. Moreover, only two studies used two distinct languages [75, 76] for dementia detection from speech. While these studies have undoubtedly contributed valuable insights into the application of machine learning in speech detection, the variability in the languages of the data is one of the challenges that could be explored further to build a system that can generalize to multiple languages. To achieve broader applicability, future studies should focus on including a more extensive array of languages. Multilingual research can offer several advantages, including the ability to develop models that are not bound by language-specific constraints, and a broader reach to a more diverse population of subjects. This expansion into multilingual research is crucial for creating inclusive and globally applicable diagnosis tools for dementia using speech analysis.

Data is also scarce, and more data need to be tested on different models [73, 74]. While the synthesized studies have made valuable contributions, the limited availability of large and representative datasets has constrained the development and generalizability of machine learning. Thus, conducting further validation for machine learning models on a larger and more diverse population is yet to be explored. Additionally, longitudinal studies are important for detecting dementia from speech, as most of the existing research relies on a cross-sectional study design, which may not capture the progressive nature of dementia.

CONCLUSION AND FUTURE DIRECTIONS

This systematic literature review has shed light on the current state of the art of dementia detection approaches using machine learning techniques through speech or gait analysis. Firstly, the review examined the machine learning algorithms used in the selected studies. Additionally, it explored the most significant features incorporated into these models. Furthermore, it analyzed the characteristics of the data employed in the selected studies. The findings of this review have shown that machine learning models hold promise to be used in the field of dementia for detection through non-invasive methods. This study shows that automatic transcription has proven to be highly effective, demonstrating commendable performance when compared to manual transcription methods. Additionally harnessing a combining linguistic, acoustic feature, and cognitive tests for speech analysis and exploring more nuanced motor changes in gait, we found them to improve the performance of the machine learning models. Moreover, dual-task has been shown to be effective in identifying individuals with dementia using machine learning. Further research to investigate the potential of employing multimodal data collection methods (combining gait and speech) to offer comprehensive assessment methods for early dementia detection is an area to be explored. This combination will provide a holistic assessment of the disease status; however, the sufficiency of the method is yet to be tested and validated.

Future direction for gait studies includes employing the multimodal data fusion method to investigate the combination of gait data with other modalities for a more comprehensive assessment and incorporating longitudinal studies to capture gait changes over time for early detection of dementia. Furthermore, the use of dual-task assessments to enhance dementia detection through gait analysis and investigating subtle motor changes in gait for more nuanced cognitive assessments can be explored, and expanding the dataset size to overcome the limitations associated with small sample sizes. Finally, utilizing machine learning applications for differentiating between var-

ious dementia subtypes including MCI, AD, vascular dementia, frontotemporal dementia, mixed dementia, and SCI.

Speech analysis may be enhanced by collecting more diverse data from various populations to enhance the generalizability of findings and utilizing deep learning techniques to merge datasets from different languages, extending models to multiple languages for wider applicability. This may include experimenting with various features, including linguistic, acoustic, and cognitive test scores, to enhance model accuracy and integrating other modalities with speech data for more comprehensive and improved predictive performance. Finally, conducting longitudinal studies to understand dynamic changes in speech for early detection.

By investigating these research avenues, we can advance the field of early dementia detection by the use of non-invasive methods, making it more accurate, effective, and robust in diagnosing this critical condition. While acknowledging the limitation of this study of focusing on the studies that are only in the English language, these future directions hold promise in addressing the multifaceted challenges of dementia diagnosis and contributing to the enhanced well-being of old people.

AUTHOR CONTRIBUTIONS

Mustafa Al-Hammadi (Conceptualization; Data curation; Formal analysis; Methodology; Visualization; Writing – original draft; Writing – review & editing); Hasan Fleyeh (Conceptualization; Methodology; Supervision; Writing – review & editing); Anna Cristina Åberg (Conceptualization; Supervision; Writing – review & editing); Kjartan Halvorsen (Conceptualization; Supervision; Writing – review & editing); Ilias Thomas (Conceptualization; Methodology; Project administration; Supervision; Validation; Writing – original draft; Writing – review & editing).

ACKNOWLEDGMENTS

The authors have no acknowledgments to report.

FUNDING

The authors have no funding to report.

CONFLICT OF INTEREST

The authors have no conflict of interest to report.

DATA AVAILABILITY

Data sharing is not applicable to this article as no datasets were generated or analyzed during this study.

REFERENCES

[1] World Health Organization. Ageing and Health. https://www.who.int/news-room/fact-sheets/detail/ageing-and-health. Accessed August 16, 2023.

[2] World Health Organization. Global Dementia Observatory (GDO). https://www.who.int/data/gho/data/themes/global-dementia-observatory-gdo. Accessed August 16, 2023.

[3] (2023) 2023 Alzheimer's disease facts and figures. *Alzheimers Dement* **19**, 1598-1695.

[4] Gauthier S, Webster C, Servaes S, Morais JA, Rosa-Neto P (2022) *World Alzheimer Report 2022. Life after diagnosis: Navigating treatment care and support*. Alzheimer's Disease International, London.

[5] Rasmussen J, Langerman H (2019) Alzheimer's disease – why we need early diagnosis. *Degener Neurol Neuromuscul Dis* **9**, 123-130.

[6] Tsoi KKF, Chan JYC, Hirai HW, Wong SYS, Kwok TCY (2015) Cognitive tests to detect dementia: A systematic review and meta-analysis. *JAMA Intern Med* **175**, 1450-1458.

[7] Haulcy R, Glass J (2021) Classifying Alzheimer's disease using audio and text-based representations of speech. *Front Psychol* **11**, 624137.

[8] Åhman HB, Berglund L, Cedervall Y, Kilander L, Giedraitis V, McKee KJ, Ingelsson M, Rosendahl E, Åberg AC (2020) Dual-task tests predict conversion to dementia—A prospective memory-clinic-based cohort study. *Int J Environ Res Public Health* **17**, 8129

[9] Nordin E, Rosendahl E, Lundin-Olsson L (2006) Timed "Up & Go" Test: Reliability in older people dependent in activities of daily living—Focus on cognitive state. *Phys Ther* **86**, 646-655.

[10] Fraser K, Fors KL, Eckerström M, Öhman F, Kokkinakis D (2019) Predicting MCI status from multimodal language data using cascaded classifiers. *Front Aging Neurosci* **11**, 205.

[11] Jamour M, Becker C, Synofzik M, Maetzler W (2012) Gangveränderungen als Frühindikator einer Demenz. *Z Gerontol Geriatr* **45**, 40-44.

[12] Windham BG, Parker SB, Zhu X, Gabriel KP, Palta P, Sullivan KJ, Parker KG, Knopman DS, Gottesman RF, Griswold ME, Mosley THJr (2022) Endurance and gait speed relationships with mild cognitive impairment and dementia. *Alzheimers Demen Diagn Assess Dis Monit* **14**, 12281.

[13] Beauchet O, Annweiler C, Callisaya ML, De Cock AM, Helbostad JL, Kressig RW, Srikanth V, Steinmetz JP, Blumen HM, Verghese J, Allali G (2016) Poor gait performance and prediction of dementia: Results from a meta-analysis. *J Am Med Dir Assoc* **17**, 482-490.

[14] Cohen J, Verghese J (2019) Gait and dementia. *Handb Clin Neurol* **167**, 419-427.

[15] Grande G, Triolo F, Nuara A, Welmer A, Fratiglioni L, Vetrano DL (2019) Measuring gait speed to better identify prodromal dementia. *Exp Gerontol* **124**, 110625.

[16] Beauchet O, Allali G, Berrut G, Hommet C, Dubost V, Assal F (2008) Gait analysis in demented subjects: Interests and perspectives. *Neuropsychiatr Dis Treat* **4**, 155-160.

[17] Ramírez F, De La Iglesia Gutiérrez M (2021) Dual-Task Gait as a predictive tool for cognitive impairment in older adults: A systematic review. *Front Aging Neurosci* **13**, 769462.

[18] Klímová B, Kuča K (2016) Speech and language impairments in dementia. *J Appl Biomed* **14**, 97-103.

[19] Szatlóczki G, Hoffmann I, Vincze V, Kálmán J, Pákáski M (2015) Speaking in Alzheimer's disease, is that an early sign? Importance of changes in language abilities in Alzheimer's disease. *Front Aging Neurosci* **7**, 195.

[20] Haider F, De La Fuente S, Luz S (2020) An assessment of paralinguistic acoustic features for detection of Alzheimer's dementia in spontaneous speech. *IEEE J Sel Top Signal Process* **14**, 272-281.

[21] Gonzalez-Atienza M, Peinado AM, Gonzalez-Lopez JA (2021) An automatic system for dementia detection using acoustic and linguistic features. *Proc IberSPEECH* **2021**, 265-269.

[22] Chiaramonte R, Cioni M (2020) Critical spatiotemporal gait parameters for individuals with dementia: A systematic review and meta-analysis. *Hong Kong Physiother J* **41**, 1-14.

[23] Buckley C, Alcock L, McArdle R, Rehman RZU, Del Din S, Mazzà C, Yarnall AJ, Rochester L (2019) The role of movement analysis in diagnosing and monitoring neurodegenerative conditions: Insights from gait and postural control. *Brain Sci* **9**, 34.

[24] Fuentes-Abolafio IJ, Stubbs B, Pérez-Belmonte LM, Bernal-López MR, Gómez-Huelgas R, Cuesta-Vargas Á (2020) Functional parameters indicative of mild cognitive impairment: A systematic review using instrumented kinematic assessment. *BMC Geriatr* **20**, 282.

[25] Mancioppi G, Fiorini L, Rovini E, Cavallo F (2021) The use of motor and cognitive Dual-Task quantitative assessment on subjects with mild cognitive impairment: A systematic review. *Mech Ageing Dev* **193**, 111393.

[26] Vigo I, Coelho L, Reis S (2022) Speech- and language-based classification of Alzheimer's disease: A systematic review. *Bioengineering* **9**, 27.

[27] De La Fuente Garcia S, Ritchie C, Luz S (2020) Artificial intelligence, speech, and language processing approaches to monitoring Alzheimer's disease: A systematic review. *J Alzheimers Dis* **78**, 1547-1574.

[28] Petti U, Baker S, Korhonen A (2020) A systematic literature review of automatic Alzheimer's disease detection from speech and language. *J Am Med Inform Assoc* **27**, 1784-1797.

[29] Yang Q, Li X, Ding X, Xu F, Ling Z (2022) Deep learning-based speech analysis for Alzheimer's disease detection: A literature review. *Alzheimers Res Ther* **14**, 186.

[30] Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group (2009) Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Ann Intern Med* **151**, 264-269.

[31] Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA (2015) Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev* **4**, 1.

[32] Xiao Y, Watson M (2017) Guidance on conducting a systematic literature review. *J Plan Educ Res* **39**, 93-112.

[33] Amir-Behghadami M, Janati A (2020) Population, Intervention, Comparison, Outcomes and Study (PICOS) design as a framework to formulate eligibility criteria in systematic reviews. *Emerg Med J* **37**, 387.

[34] Ishibashi R, Nojiri N, Saho K, Meng L (2022) Dementia diagnose based on machine learning using doppler radar image for the elderly person. *Proceedings of the 4th International Symposium on Advanced Technologies and Applications in the Internet of Things (ATAIT 2022),* Volume 3198 of CEUR Workshop Proceedings, pp. 17-24.

[35] You Z, You Z, Li Y, Zhao S, Ren H, Hu X (2021) Alzheimer's disease distinction based on gait feature analysis. *2021 IEEE International Conference on E-health Networking, Application & Services (HEALTHCOM)*.

[36] Zhang Z, Jiang Y, Cao X, Yang X, Zhu C, Li Y, Liu Y (2021) Deep learning based gait analysis for contactless dementia detection system from video camera. *2021 IEEE International Symposium on Circuits and Systems (ISCAS)*.

[37] Islam M, Chen G, Jin S (2019) An overview of neural network. *Am J Neural Netw Appl* **5**, 7-11.

[38] Chen P, Lin C, Schölkopf B (2005) A tutorial on ν-support vector machines. *Appl Stoch Models Bus Ind* **21**, 111-136.

[39] Seifallahi M, Mehraban AH, Galvin JE, Ghoraani B (2022) Alzheimer's disease detection using comprehensive analysis of timed up and go test via Kinect V.2 camera and machine learning. *IEEE Trans Neural Syst Rehabil Eng* **30**, 1589-1600.

[40] Kondragunta J, Seidel R, Hirtz G (2020) Machine learning based identification of elderly persons with cognitive impairment using dynamic time warping. *Curr Dir Biomed Eng* **6**, 360-363.

[41] Aoki K, Ngo TT, Mitsugami I, Okura F, Niwa M, Makihara Y, Yagi Y, Kazui H (2019) Early detection of lower MMSE scores in elderly based on dual-task gait. *IEEE Access* **7**, 40085-40094.

[42] LaValley MP (2008) Logistic regression. *Circulation* **117**, 2395-2399.

[43] Nielsen MS, Simonsen AH, Siersma V, Hasselbalch SG, Hoegh P (2018) The diagnostic and prognostic value of a Dual-Tasking paradigm in a memory clinic. *J Alzheimers Dis* **61**, 1189-1199.

[44] Ng TP, Lee TS, Lim WS, Chong MS, Yap P, Cheong CY, Rawtaer I, Liew TM, Gwee X, Gao Q, Yap KB (2022) Functional mobility decline and incident mild cognitive impairment and early dementia in community-dwelling older adults: The Singapore Longitudinal Ageing Study. *Age Ageing* **51**, 1-10.

[45] Sakthivadivel V, Geetha J, Gaur A, Kaliappan A (2022) Performance-Oriented Mobility Assessment test and Timed Up and Go test as predictors of falls in the elderly – A cross-sectional study. *J Fam Med Prim Care* **11**, 7294-7298.

[46] Vetráb M, Egas-López JV, Balogh R, Imre N, Hoffmann I, Tóth L, Pákaski M, Kálmán J, Gosztolya G (2022) Using spectral sequence-to-sequence autoencoders to assess mild cognitive impairment. *2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6468-6471.

[47] Themistocleous C, Eckerström M, Kokkinakis D (2018) Identification of mild cognitive impairment from speech

in Swedish using deep sequential neural networks. *Front Neurol* **9**, 975.

[48] Al-Atroshi C, Beulah JR, Singamaneni KK, Cyril CPD, Neelakandan S, Velmurugan S (2022) Automated speech based evaluation of mild cognitive impairment and Alzheimer's disease detection using with deep belief network model. *Int J Healthc Manage* **45**, 1-11.

[49] Chau HH, Chau Y, Wang HL, Chuang YF, Lee CC (2022) MCI detection based on deep learning with voice spectrogram. *2022 IEEE 4th Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS)*, pp. 212-216.

[50] Bertini F, Allevi Davide Lutero Gianluca, Montesi D, CalzàLaura (2021) Automatic speech classifier for mild cognitive impairment and early dementia. *ACM Trans Comput Healthc* **3**, 1-11.

[51] Sadeghian R, Schaffer JD, Zahorian SA (2021) Towards an automatic speech-based diagnostic test for Alzheimer's disease. *Front Comput Sci* **3**, 624594.

[52] Sadeghian R, Schaffer JD, Zahorian SA (2017) Speech processing approach for diagnosing dementia in an early stage. *INTERSPEECH 2017 – 18th Annual Conference of the International Speech Communication Association*, pp. 2706-2709.

[53] Sangchocanonta S, Vongsurakrai S, Sroykhumpa K, Ellermann V, Munthuli, A, Anansiripinyo T, Onsuwan, C, Hemrungrojn S, Kosawat K, Tantibundhit C (2021) Development of Thai picture description task for Alzheimer's screening using part-of-speech tagging. *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 2104-2109.

[54] Chien Y, Hong S, Cheah W, Yao L, Chang Y, Fu L (2019) An automatic assessment system for Alzheimer's disease based on speech using feature sequence generator and recurrent neural network. *Sci Rep* **9**, 19597.

[55] López-de-Ipiña K, Martinez-de-Lizarduy U, Calvo PM, Beitia B, García-Melero J, Fernández E, Ecay-Torres M, Faundez-Zanuy M, Sanz P (2020) On the analysis of speech and disfluencies for automatic detection of mild cognitive impairment. *Neural Comput Appl* **32**, 15761-15769.

[56] Gosztolya G, Vincze V, Tóth L, Pákáski M, Kálmán J, Hoffmann I (2019) Identifying mild cognitive impairment and mild Alzheimer's disease based on spontaneous speech using ASR and linguistic features. *Comput Speech Lang* **53**, 181-197.

[57] Yu B, Williamson JR, Mundt JC, Quatieri TF (2018) Speech-based automated cognitive impairment detection from remotely-collected cognitive test audio. *IEEE Access* **6**, 40494-40505.

[58] König A, Linz N, Tröger J, Wolters M, Alexandersson J, Robert P (2018) Fully automatic speech-based analysis of the semantic verbal fluency task. *Dement Geriatr Cogn Disord* **45**, 198-209.

[59] König A, Satt A, Sorin A, Hoory R, Derreumaux A, David R, Robert P (2018) Use of speech analyses within a mobile application for the assessment of cognitive impairment in elderly people. *Curr Alzheimer Res* **15**, 120-129.

[60] Tóth L, Hoffmann I, Gosztolya G, Vincze V, Szatlóczki G, Bánréti Z, Pákáski M, Kálmán J (2018) A speech recognition-based solution for the automatic detection of mild cognitive impairment from spontaneous speech. *Curr Alzheimer Res* **15**, 130-138.

[61] Metarugcheep S, Punyabukkana P, Wanvarie D, Hemrungrojn S, Chunharas C, Pratanwanich N (2022) Selecting the most important features for predicting mild cognitive impairment from Thai verbal fluency assessments. *Sensors* **22**, 5813.

[62] 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**, 0253988.

[63] Nagumo R, Zhang Y, Ogawa Y, Hosokawa M, Abe K, Ukeda T, Sumi S, Kurita S, Nakakubo S, Lee S, Doi T, Shimada H (2020) Automatic detection of cognitive impairments through acoustic analysis of speech. *Curr Alzheimer Res* **17**, 60-68.

[64] Munthuli A, Vongsurakrai S, Anansiripinyo T, Ellermann V, Sroykhumpa K, Onsuwan C, Chutichetpong P, Hemrungrojn S, Kosawat K, Tantibundhit C (2021) Thammasat-NECTEC-Chula's Thai language and cognition assessment (TLCA): The Thai Alzheimer's and Mild Cognitive Impairment Screening Test. *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 690-694.

[65] Eyben F, Wöllmer M, Schuller B (2010) OpenSMILE – The Munich versatile and fast open-source audio feature extractor. *MM'10 – Proceedings of the ACM Multimedia 2010 International Conference*, pp. 1459-1462.

[66] Cunningham P, Delany SJ (2021) K-Nearest Neighbour Classifiers – a tutorial. *ACM Comput Surv* **54**, 1-25.

[67] Mirzaei S, El Yacoubi M, Garcia-Salicetti S, Boudy J, Kahindo C, Cristancho-Lacroix V, Kerhervé H, Rigaud AS (2018) Two-stage feature selection of voice parameters for early Alzheimer's disease prediction. *IRBM* **39**, 430-435.

[68] Nasrolahzadeh M, Mohammadpoory Z, Haddadnia J (2018) Higher-order spectral analysis of spontaneous speech signals in Alzheimer's disease. *Cogn Neurodyn* **12**, 583-596.

[69] Sagi O, Rokach L (2018) Ensemble learning: A survey. *WIREs Data Min Knowl Discov* **8**, e1249.

[70] You Y, Ahmed B, Barr P, Ballard K, Valenzuela M (2019) Predicting dementia risk using paralinguistic and memory test features with machine learning models. *2019 IEEE Healthcare Innovations and Point of Care Technologies (HI-POCT).*

[71] Weiner J, Frankenberg C, Schröder J, Schultz T (2019) Speech reveals future risk of developing dementia: Predictive dementia screening from biographic interviews. *2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU).*

[72] De Looze C, Kelly F, Crosby L, Vourdanou A, Coen RF, Walsh C, Lawlor BA, Reilly RB (2018) Changes in speech chunking in reading aloud is a marker of mild cognitive impairment and mild-to-moderate Alzheimer's disease. *Curr Alzheimer Res* **15**, 1-20.

[73] Soroski T, Vasco T da C, Newton-Mason S, Granby S, Lewis C, Harisinghani A, Rizzo M, Conati C, Murray G, Carenini G, Field TS, Jang H (2022) Evaluating web-based automatic transcription for Alzheimer speech data: Transcript comparison and machine learning analysis. *JMIR Aging* **5**, e33460.

[74] Liang X, Batsis JA, Zhu Y, Driesse T, Roth RM, Kotz D, MacWhinney B (2022) Evaluating voice-assistant commands for dementia detection. *Comput Speech Lang* **72**, 101297.

[75] Kálmán J, Devanand DP, Gosztolya G, Balogh R, Imre N, Tóth L, Hoffmann I, Kovács I, Vincze V, Pákáski M (2022) Temporal speech parameters detect mild cognitive impairment in different languages: Validation and comparison of the Speech-GAP Test® in English and Hungarian. *Curr Alzheimer Res* **19**, 373-386.

[76] Schäfer S, Mallick E, Schweda L, König A, Zhao J, Linz N, Bodin TH, Skoog J, Possemis N, Huurne DT, Zettergren A, Kern S, Sacuiu S, Ramakers I, Skoog I, Tröger J (2023) Screening for mild cognitive impairment using a machine learning classifier and the remote speech biomarker for cognition: Evidence from two clinically relevant cohorts. *J Alzheimers Dis* **91**, 1165-1171.

[77] Reeves SM, Williams VJ, Blacker D, Woods RL (2023) Further evaluation of narrative description as a measure of cognitive function in Alzheimer's disease. *Neuropsychology* **37**, 801-812.

[78] Li RA, Hajjar I, Goldstein F, Choi JD (2020) Analysis of hierarchical multi-content text classification model on B-SHARP dataset for early detection of Alzheimer's disease. *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, pp. 358-365.

[79] López-de-Ipiña K, Martinez-de-Lizarduy U, Calvo PM, Mekyska J, Beitia B, Barroso N, Estanga A, Tainta M, Ecay-Torres M (2018) Advances on automatic speech analysis for early detection of Alzheimer disease: A non-linear multi-task approach. *Curr Alzheimer Res* **15**, 139-148.

[80] Praat: Doing phonetics by computer. https://www.fon.hum.uva.nl/praat/. Accessed November 21, 2023.

[81] Chiche A, Yitagesu B (2022) Part of speech tagging: A systematic review of deep learning and machine learning approaches. *J Big Data* **9**, 10.

[82] Robinson L, Tang E, Taylor J (2015) Dementia: Timely diagnosis and early intervention. *BMJ* **350**, h3029.