Simulation Modelling Practice and Theory 99 (2020) 102023

Contents lists available at ScienceDirect

Simulation Modelling Practice and Theory

journal homepage: www.elsevier.com/locate/simpat

A new machine learning method for identifying Alzheimer's disease

Lin Liu[a], Shenghui Zhao[b,*], Haibao Chen[b], Aiguo Wang[b]

[a] School of Computer Science and Engineering, Anhui University of Science and Technology, Huainan, Anhui, China
[b] School of Computer and Information Engineering, Chuzhou University, Chuzhou, Anhui, China

ARTICLE INFO

Keywords:
Alzheimer's disease
Machine learning
Spectrogram
Speech recognition

ABSTRACT

Most of the studies on Alzheimer's disease (AD) have been carried out using medical images. However, the acquisition of medical images data is difficult. The identification based on the patient's speech data can effectively reduce the medical cost, and the speech data can be collected in a non-invasive manner so that the patient's data can be collected in real-time and accurately. This paper proposes a new method that uses the spectrogram features extracted from speech data to identify AD, which can help families to understand the disease development of patients in an earlier stage, so that they can take measures in advance to delay the disease development. We use the speech data collected from the elderly that express the speech features displayed in the speech and used the machine learning methods for identifying AD. During the simulation and experiment, we collect a new speech dataset, which includes Alzheimer's disease patients and healthy control subjects. Then, we compare with the speech data made available by the Dem@Care project. Among the tested models, LogisticgressionCV model exhibited the best performance. It is shown that this method using extracted spectrogram features from speech data to identify AD is feasible. The credibility of the new dataset and feasibility of the used methods in this paper are demonstrated.

1. Introduction

There are many types of dementia, including vascular dementia (VD), Lewy body dementia (LBD), and especially Alzheimer's disease (AD), which is a neurodegenerative disease that commonly occurs in the elderly persons. According to the World Alzheimer's Disease 2018 report [1], a patient with dementia will be born every 3 s in the world. In 2018, about 50 million people all over the world suffered from dementia. By 2050, the number will increase to 152 million, which is three times that of the present number of the dementia. The number of AD accounts for 60% – 80% of dementia and is a major burden on the physical and mental condition of the elderly. The theme of World Alzheimer's Disease Day on September 21, 2019, is "Memory 3 S", which is designed to remind us that every Alzheimer's disease patient will appear every 3 s around the world. The report also shows that between 2010 and 2050 the number of dementia patients mostly occur in low and middle-income countries, and it increases faster and in larger numbers compared to that in the high-income countries.

Patients with Alzheimer's disease may suffer from memory loss, gradual impairment of language function, unresponsiveness, gradually decline in daily living ability and unusual changes in personality and behavior. More seriously, they do not remember friend and even their loved ones. It also brings a great mental burden to both themselves and their families. Alzheimer's disease is more of a mental blow and irreversible. All treatments can only delay the process of deterioration and cause patients and their families to collapse. The study in [2] found that caregivers of AD had a higher incidence of depression. It shows the necessity of early

* Corresponding author.
E-mail addresses: lingting@chuzu.edu.cn (L. Liu), shzhao@ah.edu.cn (S. Zhao), chb@chzu.edu.cn (H. Chen).

https://doi.org/10.1016/j.simpat.2019.102023
Received 16 October 2019; Received in revised form 4 November 2019; Accepted 10 November 2019
Available online 12 November 2019
1569-190X/

prediction, identification, and intervention measures for AD. Although this disease is irreversible, we can identify the disease in its early stage and take intervention measures on environment and drugs to slow down the progression of the disease.

Most of the existing studies on Alzheimer's disease are based on medical images. With the development of artificial intelligence, computer-aided diagnosis, and classification of AD, and prediction of AD transformation process become a research hotspot in recent years. To obtain medical images, patients are required to go to hospital for physical check. The costs are too high for low and middle families to do hospital check regularly. Therefore, the diagnosis of the AD can be delayed.

Language disorder is also one of the symptoms of AD. In the process of speaking, AD will cause memory barriers, and cause difficulties in finding words, resulting unclear speech, and pauses. This paper is based on language processing to study the speech feature of AD.

In our study, we used IoT devices and collected 23 elderly persons' speech data and then utilized machine learning methods to identify AD and healthy control (HC) groups. That is, combining AI technology with the voice of subjects to detect the subtle changes that cannot be heard from human ears.

In 1906, a German neuropathologist Alzheimer first published a case of AD [3], which was named Alzheimer's disease in 1910. However, the true causes of this disease have not been clearly identified until now.

Most of the current studies on AD is based on artificial intelligence through using the Alzheimer's Neuroimaging Initiative (ADNI) dataset. ADNI dataset includes magnetic resonance imaging (MRI) and positron emission tomography images (PET), genetics, cognitive tests, cerebrospinal fluid, and blood biomarkers as predictors of the disease. Studies in [4, 6, 8, 9] used the ADNI dataset including MRI and PET images to classify or identify AD experiments. They applied the Deep Neural Network (DNN), Sparse Classifier (SRC), Random Forest (RF) and Convolutional neural network (CNN), respectively. The study in [5] proposed a new algorithm that is MM-SDPN algorithm to diagnosis AD, which is based on MRI and PET images data. The study in [10] proposed a method that used CNN to study the subjects' brain images with Computed Tomography (CT) for classifying AD. The study [11] proposed using brain waves (EEG) for classification AD with CNN. The study in [12] proposed using classical and deep learning technology for the diagnosis and detection of AD. The study in [13] used a parameter-based deep learning method to predict the transition of mild cognitive impairment MCI to Alzheimer's disease AD. Next, we introduce the related studies with speech aspects.

The study in [14] claimed that using language analysis methods was more sensitive than other cognitive tests for detecting AD. Temporal characteristics of spontaneous speech, such as speech speed, number of speech pauses, and duration of their speech time length, were sensitive detectors in the early stage of the disease, enabling early simple language screening for Alzheimer's disease. The study in [15] proposed the potential of language tasks in the identification of Alzheimer's disease, with the aim of exploring the diagnostic effect of language tasks on Alzheimer's disease. The study in [16] used linguistic features to identify Alzheimer's disease in the narrative discourse, which aimed to show the latest accuracy of the automatic identification of Alzheimer's disease through a short narrative sample from the picture description task, and its final identification accuracy is 81%. The study in [17] also used the method based on image description to diagnose each group of 18 subjects, achieving an accuracy of 75%.

Table 1 summarizes the above research literature related to AD. It mainly includes the data used in the studies, the methods used and the results achieved.

Table 1
Summary of the related works on AD.

Studies	Data	Methods used	Results
[4]	MRI and FDG-PET	Deep-learning-based framework	MCI Identifying accuracy 82.4%. Probable AD Classifying sensitivity of 94.23%. HC Classifying specificity 86.3%.
[5]	MRI and PET	MM-SDPN algorithm	Proposed MM-SDPN can effectively learn and fuse multimodal data for the diagnosis of AD.
[6]	MRI and PET	Spare features of data and SVC	HC VS AD classification accuracy 92%. HC VS MCI classification accuracy 84%.
[7]	sMRI, FDG-PET and florbetapir PET	Sparse inverse covariance estimation	The connections among different brain regions were quite different between HC/MCI and MCI/AD.
[8,9]	fMRI	Random Forest [8], CNN [9]	HC VS AD classification sensitivity 84% and specificity 92.3% [8]. AD Classification accuracy of 96.86%[9].
[10]	CT	CNN	Average Classification accuracy 86.8%.
[11]	EEG	CNN	Average Classification accuracy 80%.
[12]	MRI	DNN, SVM	Classification accuracy 95.93% on OASIS dataset. Classification accuracy 98.74% on ADNI dataset.
[13]	Demographic, neuropsychological, and APOe4 genetic data	Deep learning architecture	Prediction AUC 0.925, accuracy 86%, sensitivity 87.5%, specificity 85%.
[14]	Speech data of Language tasks	Statistical analysis	Language deficit in AD is present in the early stage of the disease.
[15]	Speech data of Language tasks	Discriminant analysis	The Sentence Correction Task, the MSQ, and the Verbal Expression Test to best discriminate patients.
[16]	Picture description task speech data of DementiaBank	Machine learning	Classification accuracy above 81%.
[17]	Transcripts of connected speech	Machine learning	Classification highest accuracy 75%.

Medical images have been widely used in computer-aided research and have achieved a great success. For magnetic resonance images (MRI), there exist many studies. However, these images vary due to different types of devices. The medical images obtained from one device is better diagnosed by AI technology, but it may not be effective in others of images diagnosis.

Speech processing and machine learning not only applied in medical images and natural language processing but in natural disaster resilience [18], extracting feature for structured representation of text data [19], identifying sleep disordered breathing [20], color-based object segmentation method using artificial neural network [21], improving overall process at the cardiac catheterization lab [22].

Based on the speech features to identify AD and most of the speech data based on the picture description task or memory description task for identification and diagnosis. Then extract the numerical parameter characteristics of a person's speech data, including speech rate, number of pauses, reaction time, etc., which are more limited in the size of the dataset. Therefore, this paper proposes a new method that is based on the features of the extracted speech data to identify Alzheimer's disease.

The main contributions of this paper are summarised as follows:

- a new machine learning method for identifying AD.
- collected a new dataset that was about the speech data of the elderly (VBSD).
- an algorithm to extract spectrogram automaticaly based on the audio data (Algorithm 1).
- an algorithm to identify AD based on machine learning (Algorithm 2).

The rest of the paper is organised as follows. In Section 2, we present the methodology. In Section 3, we present the experiment process and result discussions.

2. Methodology

The system consists of a wearable IoT device that records the person's voice all the time [23]. The voice data is transmitted to the cloud server, where the original speech data is stored. The receiving new data will be recognized using the trained model for identifying the AD symptoms, as shown in Fig. 1. Firstly, we collect the original audio data of all the subjects. The audio data can be divided different number of audio segments due to the person cannot speak all the time which must contain the silence audio segments. Secondly, dividing each audio segment into more multi-segments with the duration of 1 s. Thirdly, extracting the spectrogram features from the audio segments. The obtained spectrogram data can be used to training model [24]. Finally, using this trained model to identify a set of new data whether it is belongs to AD.

2.1. Identifying AD based on speech data

This paper focus on the spectrogram features of speech data to identify AD. The machine learning techniques that currently widely used in varies domains are used. In the next section, the meaning of the spectrogram is described briefly and details of data collection and processing are as follows.

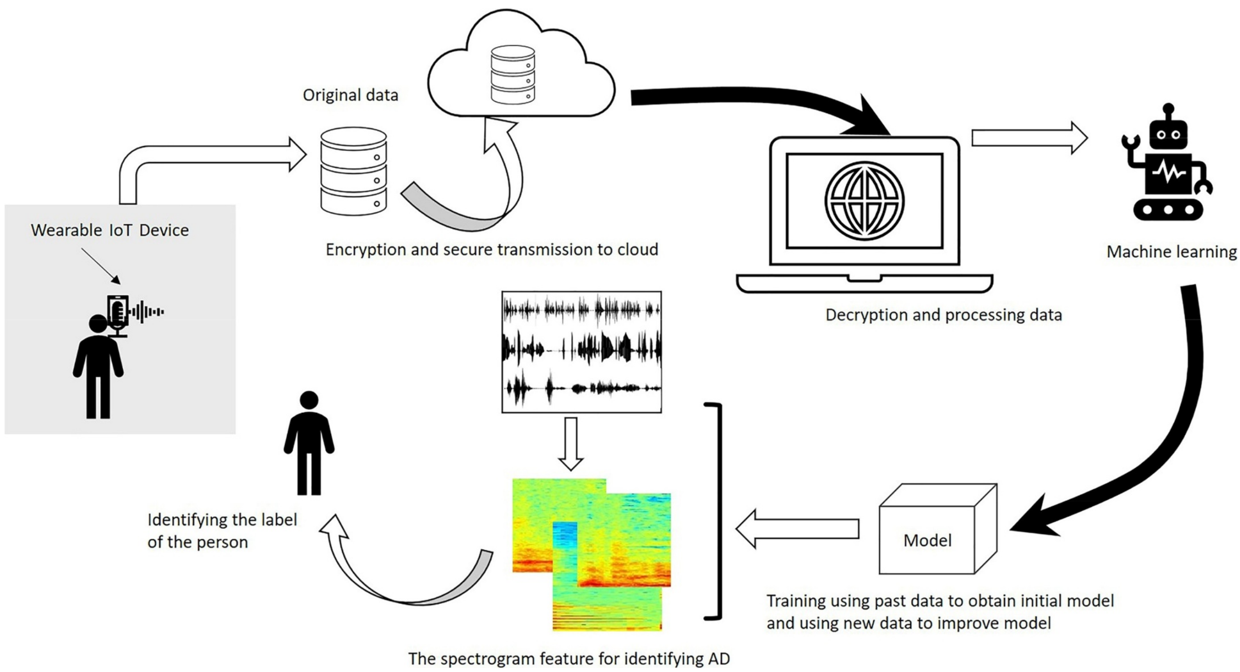

Fig. 1. The overall system architecture.

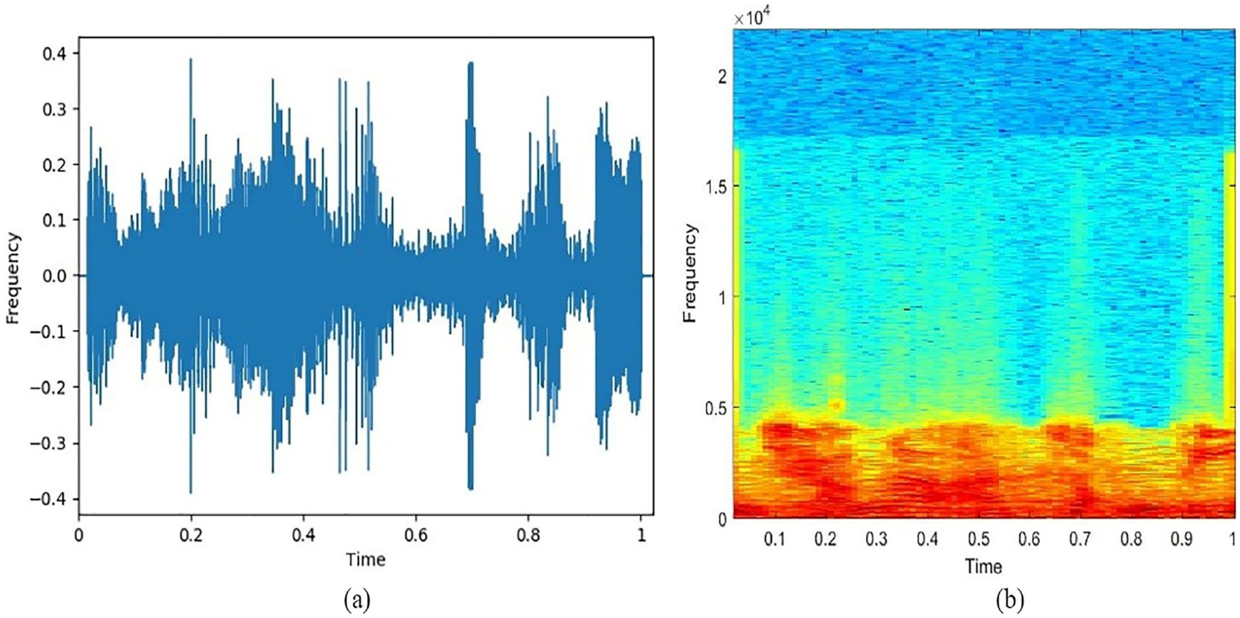

Fig. 2. (a) The HC original speech waveform. (b) The HC spectrogram.

2.1.1. Introduction to the spectrogram

The spectrogram is a representation of a time-frequency signal that represents the relationship between the time domain and the frequency domain in the form of a two-dimensional map. Different features in speech have different manifestations in the spectrogram. Therefore, compared with extracting multiple parameter features in speech signals, there are also great advantages in using a spectrogram to explore speech features, such as greatly reducing computational complexity.

After the collected speech data is unified in format, the features of the speech data are extracted according to the function of extracting the speech spectrogram. Finally, we obtained the spectrogram in the form of Fig. 2(b) or Fig. 3(b). From Fig. 2(b) and 3(b) below, we can observe the signal strength of different frequency bands over time. We can also see a strip of horizontal stripes, called "Voiceprint". The stripe is actually the place where the deep color points gather. When it continues with time, it is extended into stripes, which means that the frequency value in the speech is stronger than the horizontal coordinate value of the point, and the proportion in the whole voice is significant.

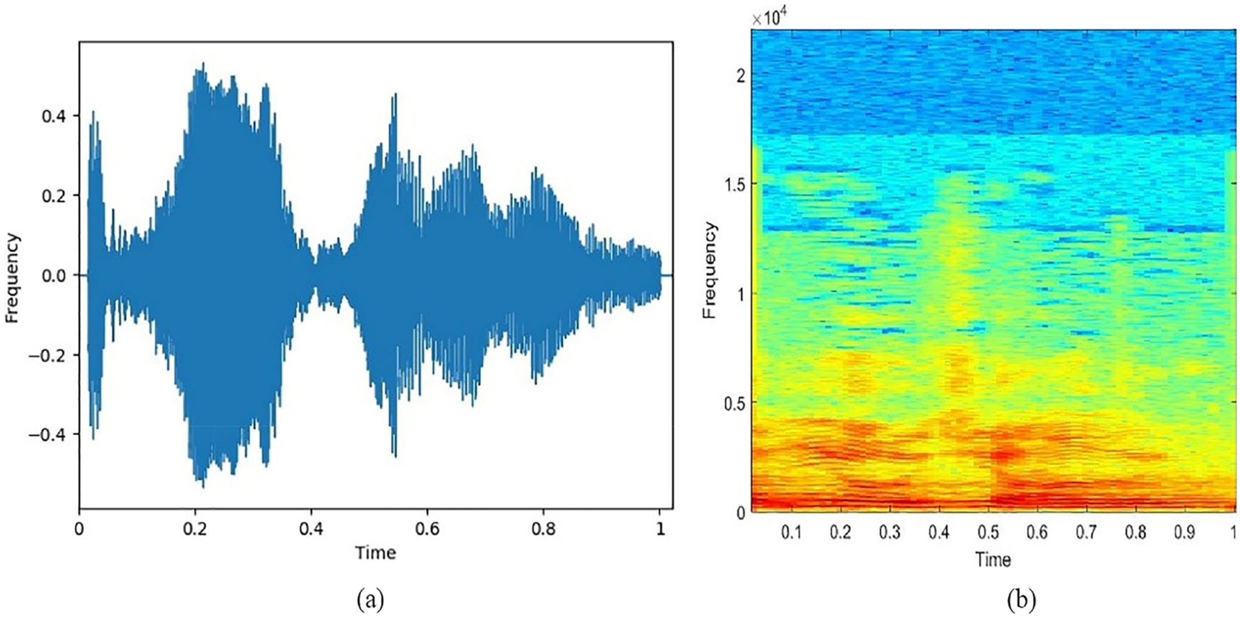

Fig. 3. (a) The AD original speech waveform. (b) The AD spectrogram.

Fig. 2(a) below is the original waveform of a normal person, from which only the subject's speech fluctuations can be seen, and Fig. 2(b) shows that the spectrogram feature extracted from the speech segment, the energy situation of a person can be seen. This paper is based on the characteristics of the spectrogram to identify AD. Fig. 3(a) is a speech waveform diagram of an elderly person suffering from AD, and Fig. 3(b) is the AD's spectrogram feature.

The horizontal axis of the spectrogram represents time and the vertical axis represents frequency. The coordinate point value is voice data energy. Since the two-dimensional plane is used to express three-dimensional information, the magnitude of the energy value is represented by color, and the darker the color is, the stronger the speech energy is. The effect of affecting people's perception is much stronger. From this observation, the artificial intelligence machine learning for the purpose of identifying the AD.

2.2. Collection and processing of speech dataset

This subsection mainly describes the procedure of data collection and processing. The formulation express and algorithm implementation are shown as follows.

2.2.1. Problem description

This section aims at solving the problem of speech data collection and processing. The collected original speech data comes from the wearable IoT device, data processing is based on spectrogram function. In order to facilitate the description of the problem, the following formulation is defined:

2.2.1.1. Data collection. Let $\mathbb{A} = \{A_1, A_2, ..., A_n\}$ denote the audio set of all subjects and $\mathbb{S} = \{S_1, S_2, ..., S_n\}$ denotes an individual of subjects. That is, S_i denotes the ith person, A_i denotes the audio data of the ith person. That is,

$$\forall A_i \in S_i, \text{ and } 0 < i \leq n \tag{1}$$

where n denotes the number of subjects that took part in the experiment. Given an audio data A_i that belongs to its subjects and i must obey the constraint of (1). Each A_i also including multi-audio segments which can be expressed by notation $A_i = \{D_1, D_2, ..., D_m\}$ and it needs obey the sample balance constraint (2). That is,

$$\forall\ Group_i\{D_1, D_2, ..., D_m\} \in A_i, \text{ and } 0 < i \leq n, \text{ and } \sum_{i=1}^{m} D_i^{time} \geq 2s \tag{2}$$

where $Group_i\{D_1, D_2, ..., D_m\}$ denote a set of audio segment of A_i, D_i^{time} denote the total duration of the ith subject. The values of m are not fixed that is because the audio segment of each subject is not equal. But in order to ensure the balance of samples, D_i^{time} is limited by (2).

2.2.1.2. Data segmentation. Dividing the obtained $A_i = \{D_1, D_2, ..., D_m\}$ into $D_j^{A_i} = \{d_1, d_2, ..., d_p\}$ that is because the duration of the audio segment is different so that needs standardize the duration time for 1 s. $D_j^{A_i}$ denotes the subsegment of the jth audio segment of A_i. That is,

$$\forall\ Group_j\{d_1, d_2, ..., d_p\} \in D_j, \text{ and } \sum_{j=1}^{m}\sum_{i=1}^{p} d_i \geq 2, \text{ and } 0 < j \leq m \tag{3}$$

where $Group_j\{d_1, d_2, ..., d_p\}$ denotes a subset of the set of audio segment, $\sum_{j=1}^{m}\sum_{i=1}^{p} d_i$ means that the total number of audio segments of a person. The values of p are not fixed that also because the duration of audio segments is not equal.

2.2.1.3. Extract spectrogram features. Extracting the spectrogram features from $D_j^{A_i} = \{d_1, d_2, ..., d_p\}$ and then using the function of spectrogram to extract all audio segments of each subject. Let $S_j^{A_i} = \{\mathcal{S}_1, \mathcal{S}_2, ..., \mathcal{S}_p\}$ denote the spectrogram features from the jth audio segment of the ith person's audio data. That is,

$$S_j^{A_i} = \{\mathcal{S}_1, \mathcal{S}_2, ..., \mathcal{S}_p\} = \{function(d_1), function(d_2), ..., function(d_p)\} \tag{4}$$

where $function(\bullet)$ denotes the function of extract spectrogram features. Let S^* denote all spectrogram features dataset of all subjects, and $S^* = \sum_{i=1}^{n}\sum_{j=1}^{m} S_j^{A_i} = \{S_1, S_2, ..., S_n\}$.

2.2.1.4. Divide training sets and testing sets. Dividing all data into training sets and testing sets is needed before simulation and experiment with machine learning methods. Fig. 4 shows that the procedure of data divided, which is similar to k fold cross-validation. However, the difference is that there are multi-data of testing sets in this experiment. The values of r denote the number of spectrogram features of a person.

The method can also be called leave one cross-validation. The leave one method not only has one data but multi-data, which changes follows S_i. Finally, the results of each metric returned to integrate the results of the simulation and experiment results for all subjects. Let $S^* = \{S_1, S_2, ..., S_n\}$ denote the all spectrogram features of all subjects that take part in this experiment, where $S_i = \{\mathcal{S}_1, \mathcal{S}_2, ..., \mathcal{S}_r\}$.

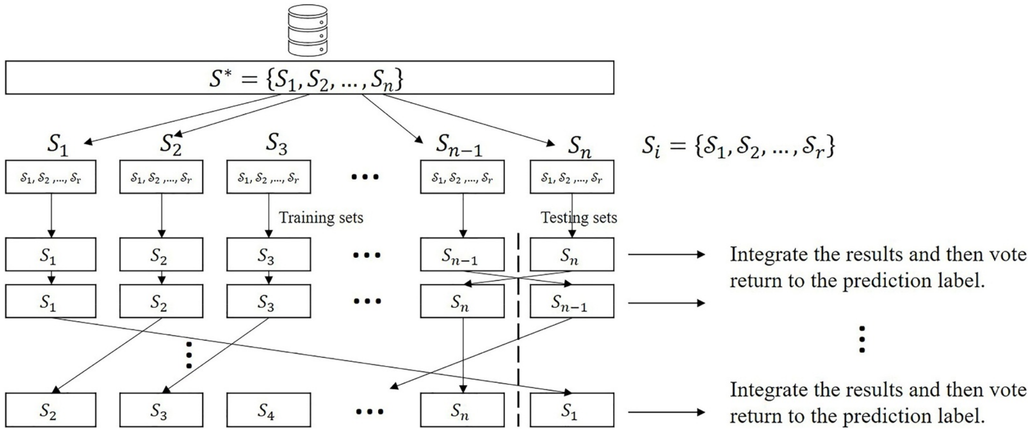

Fig. 4. Schematic diagram of data division.

$$\begin{cases} \sum_{i=1}^{r} \mathcal{S}_i \geq r - \sum_{i=1}^{r} \mathcal{S}_i, & y = 1; \\ \sum_{i=1}^{r} \mathcal{S}_i < r - \sum_{i=1}^{r} \mathcal{S}_i, & y = 0. \end{cases} \tag{5}$$

where $\mathcal{S}_i$ denotes the prediction labels that according to the i th spectrogram, and the result expressed by $y = 0/1$. $\sum_{i=1}^{r} \mathcal{S}_i$ denote the total sum of identifying data as 1 and $r - \sum_{i=1}^{r} \mathcal{S}_i$ denote the total sum of identifying data as 0. Compare the two values and obtain the final label of S_i. Then the metrics performance and labels of all data have obtained.

2.2.1.5. Machine learning models. We now discuss the simulation and experiment with machine learning methods. We use $\mathcal{M} = \{M_1, M_2, \ldots, M_k\}$ to denote all models of machine learning and $\mathcal{R} = \{X_1, X_2, \ldots, X_n\}$ to denote the set of identification results of a model, where

$$X_i = M_j(S_i), \quad 0 < i \leq n \text{ and } 0 < j \leq k \tag{6}$$

All the above stages are processed step by step. Fig. 5 shows the formulation representation, where each level has detailed stages.

2.2.2. Algorithm implementation

Algorithm 1 in Table 2 completes the extraction procedure of spectrogram features. First, it collects the audio data of all subjects from wearable device. Second, it divides the audio segment to facilite extracting spectrogram features. Finally, it inputs the audio segment and outputs the spectrogram features. In the process of implementing the algorithm, the function of extracting the spectrogram is used to generate the required data automatically.

Algorithm 1 in Table 2 that extracts spectrogram from the audio data is based on the following input and output.

Input: $D_j^{A_i} = \{d_1, d_2, \ldots, d_p\}$

Output: $S_j^{V_i} = \{\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_p\}$

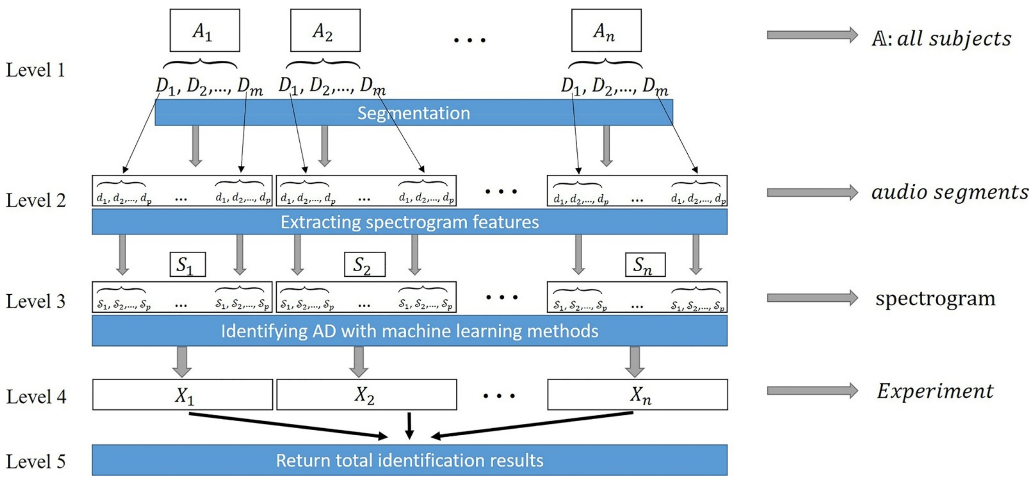

Fig. 5. The procedure of data formulation processing.

Table 2
The algorithm of data collecting and processing.

Algorithm 1 Extract spectrogram based on the audio data

Collect audio data of all subjects from wearable device
for i in range(1, $n+1$):
if the total duration of a person's audio <2 s:
delete this audio
else:
put this audio into A_i
Divided audio dataset A_i into the multi-audio segment as $D_1, D_2, \ldots, D_m$
Divided the $A_i = \{D_1, D_2, \ldots, D_m\}$ into $D_j^{A_i} = \{d_1, d_2, \ldots, d_p\}$
Input:audio segments of all subjects $D_j^{A_i} = \{d_1, d_2, \ldots, d_p\}$
for i in range(1, $n+1$):
for j in range(1, $m+1$) :
extract spectrogram feature $S_j^{V_i} = \{\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_p\}$
$= \{function(a_1), function(a_2),$
$\ldots function(a_p)\}$
Output: $S_j^{V_i} = \{\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_p\}$

Algorithm 2 in Table 3 describes the process of all machine learning models $\mathcal{M} = \{M_1, M_2, \ldots, M_k\}$ that identify AD by training sets and testing sets. First, k different machine learning models are selected, and then the models identify the labels of all subjects according to the input spectrogram features. Finally, the performance of each model is evaluated according to the evaluation metrics.

Algorithm 2 in Table 3 that performs machine learning models identifying AD is based on the following input and output.

Input: $S = \{S_1, S_2, \ldots, S_n\}$

Output: 0/1 (0 represents HC and 1 represents AD)

3. Experiment process and result discussions

This section focus on the details of the simulation and experiment procedure and their results. In addition to a brief introduction to the two datasets used, we also describe the evaluation metrics, experiment procedure and the results obtained.

3.1. Dataset description

In order to verify the feasibility of the proposed method in this paper, simulation and experiment were conducted on two datasets: One is a new dataset that we collected and we can call it VBSD dataset. The other is the Dem@Care dataset, which is made available by the Dem@Care Lab project. Through comparing the two datasets, we also verify the reliability of the collected dataset in this paper. The detail of the two datasets will be described as follows.

3.1.1. VBSD dataset

The original dataset VBSD that we collected comes from the wearable IoT devices. The obtained final spectrogram dataset can be accessed on a repository Github (https://github.com/LinLLiu/AD). The sampling frequency of each audio is set to 44.1KHZ, and the duration is set to 1 s. Then, we extract the spectrogram features from audio data, ans we feed them into machine learning classifier models in our simulation and experiment. A total of 36 subjects were included in the collected speech dataset. In order to find the common features of multiple AD patients, speech data of multiple AD patients for training. We extracted 254 AD speech data and 250

Table 3
The algorithm of machine learning models to identify AD.

Algorithm 2 Machine learning models identifying AD

Select machine learning methods $\mathcal{M} = \{M_1, M_2, \ldots, M_k\}$
Input: spectrogram features of all subjects $S = \{S_1, S_2, \ldots, S_n\}$
for i in range(1, $k+1$):
for j in range(1, $n+1$):
return the results $\mathcal{R} = \{X_1, X_2, \ldots, X_n\}$
if $\sum_{i=1}^{r} \mathcal{S}_i \geq r - \sum_{i=1}^{r} \mathcal{S}_i$:
print(model M_i identifying result with 1)
else:
print(model M_i identifying result with 0)
Output: 0/1

Table 4
The VBSD dataset used in the simulation experiment.

Group	Age range	Males	Females
AD	65–94	10	13
HC	65–92	5	8

HC speech data from data processing steps in the previous introduction. We obtained a total of 504 speech data, which means that 504 spectrogram features were extracted.

Table 4 shows the data collected in this study. Among them, the age of AD was between 65–94 years old. A total of 23 speech data of elderly people with Alzheimer's Disease (AD) were collected, and 254 spectrogram features were extracted from them. Through healthy control (HC) of elderly people between the ages of 65–92, we collected a total of 13 normal elderly speech data, extracted 250 spectrogram features. The purpose of this study is to use the speech data of elderly people with AD and HC to find the universal characteristics of all AD 's speech patterns to identify AD.

3.1.2. Dem@Care dataset

The second dataset of speech data is made available by the Dem@Care [25] project, which was collected in their laboratory and in the participant's family. The data we used include 24 speech data from AD patients and 8 speech data from HC subjects. Dem@Care related dataset literature including [26–28].

Dem@Care's speech data was collected targeted on four tasks.

- Task 1 asks an elderly person to describe a picture while looking at it;
- Task 2 asks an elderly person to describe a picture from memory after having looked at a picture;
- Task 3 asks an elderly person to follow the experiment supervisor to repeat a set of sentences;
- Task 4 asks an elderly person to repeat the pronunciation "pa-ka-ka" as soon as possible.

The studies in Aharon Satt [26, 28] used the dataset to extract the parameter features of speech data, including the number of pauses, the reaction time, and the speech rate, etc., to establish the connection between AD and its voice. These studies also found that Task 3 and Task 4 did not perform well. Therefore, for this reason, these two speech tasks were not selected as experimental data in our study.

In our simulation experiment, we applied the same configuration, settings and processing method for the VBSD dataset and Dem@Care dataset. We perform the same data processing based on all subjects to extract the spectrogram features. When speech data processing is performed, each speech duration is set to 1 s, and the sampling frequency is reset to 44.1 kHz. AD splits 257 speech segments, HC splits 231 speech segments. We finally obtains 488 spectrogram features. We use these 488 features for our experiment.

3.2. Simulation experiment

The simulation experiment details, including performance metrics, the procedure of each stage, and the results obtained, are discussed in the following subsections.

3.2.1. Performance metrics

In this paper, four metrics, *Accuracy, Precision, Recall*, and *F*1-*Score* are used for verifying the final performance of all methods used.

- *Accuracy* represents the overall identification accuracy;
- *Precision* represents the ratio of people who are predicted to be sick and their true status is sick to the number of all the subjects;
- *Recall* is the ratio of the number of people who are predicted to be sick and their true status is sick to the number of all the AD patients.
- *F*1-*Score* is Precision and Recall's weighted harmonic meaning, the higher the F1-Score, the more effective of the model.

Table 5 shows the confusion matrix of the classification results. The calculation method of each metric is given in Exp (7), where 1 represents AD, and 0 represents HC.

Table 5
Classification result confusion matrix.

Actual Class	Predicted Class	
	1	0
1	TP(True Positive)	FN(False Negative)
0	FP(False Positive)	TN(True Negative)

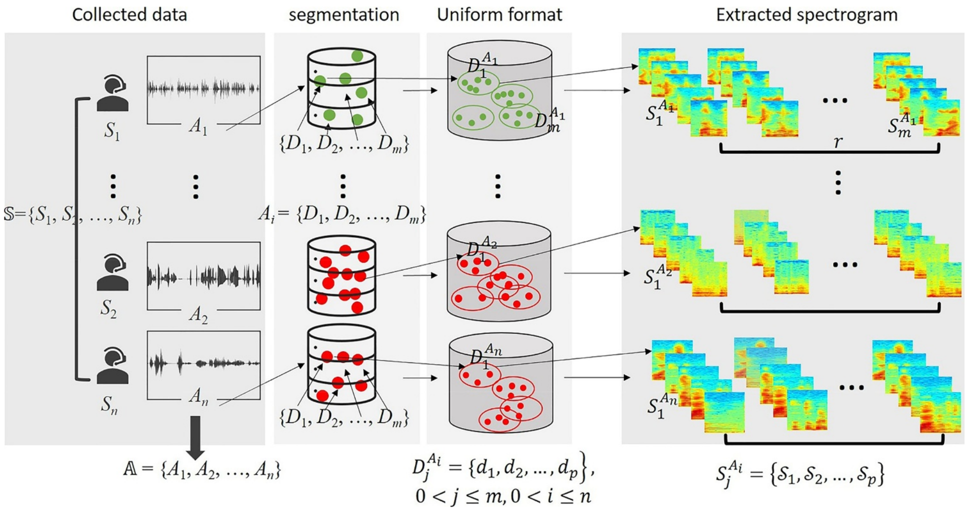

Fig. 6. The produre of extracting spectrogram.

$$\begin{cases} Accuracy = \frac{TP+TN}{TP+FN+FP+TN}; \\ Precision = \frac{TP}{TP+FP}; \\ Recall = \frac{TP}{TP+FN}; \\ F1-Score = \frac{2*Precision*Recall}{Precision+Recall}. \end{cases} \tag{7}$$

3.2.2. Simulation and experimental procedure

Before dividing (data) training sets and testing sets, we used the Yellowbrick visual diagnostic tool for assessing the scores of each model, in order to find the machine learning models that are suitable for the experimental data. We finally selected the models of LogisticRegressionCV, LinearSVC, DecisionTreeClassifier, BaggingClassifier, and MLPClassifier for simulation and experiment, because they have higher ranking scores.

Considering the dimensionality reduction parameters of Principal Component Analysis (PCA) may produce an impact on the machine learning model, we also conducted simulation on different parameter values to find better parameter values that could achieve the best performance. Fig. 6 is the produre of extracting spectrogram. We first collect the speech data of all subjects and divide the original speech data into the n-segment speech of equal size, and we extract n spectrograms from the n-segment speech data. In this way, multiple sample data for a single person are obtained. Then, based on the data, machine learning methods are used in simulation experiments to identify whether a person is AD patient.

Fig. 7 shows that the produre of machine learning models identifying AD. When executing identification experiment, it is necessary to extract the multi-samples data of the single person as the testing set, and the testing set belongs to the speech data of the same subject. It means that the training sets contain n-1 subjects and the data of testing set belongs to the nth subject. The training sets and the testing sets are mutually exclusive, that is, a person's data cannot appear in the training set and the test set at the same time. Thus, the category with the most data labels in the integration testing set is the category to which the testing set subjects belong to. The rest of the data samples are used for training.

3.3. Simulation experiment results

During the simulation experiment, each model performed 68 times in the simulation experiment according to different testing sets, where VBSD dataset is used 36 times and Dem@Care is used 32 times. Each experimental result is calculated based on the integration of all testing set identification results. Fig. 8 shows that the visualization of the experimental results of various metrics and models. From the experimental results, it can be seen that the LogisticRegressionCV, LinearSVC and MLP models perform well both in the VBSD dataset and the Dem@Care dataset.

Tables 6 and 7 show the details of the simulation experiment results of different machine learning methods based on the VBSD dataset and the Dem@Care dataset.

From the experimental results, LogisticRegressionCV obtained the best performance compared with other classification methods. Among them, F1-Score reached 86.9% on the VBSD dataset and 89.4% on the Dem@Care dataset. LinearSVC and MLP models perform relatively well that their accuracy reached above 70%. However, DecisionTree and Bagging obtained poor performance because DecisionTree is so easily producing a complex model that makes the generalization ability perform not well. Moreover, it also has the disadvantage of instability. A small change of data may lead to different tree generation. In this simulation experiment, the

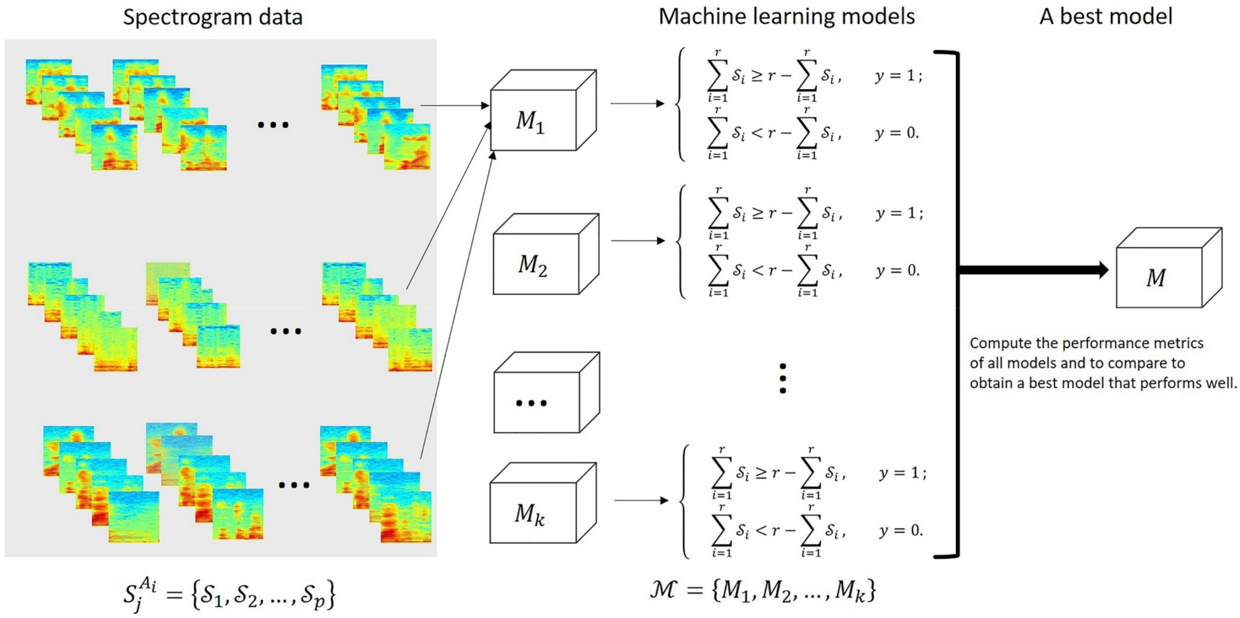

Fig. 7. The produre of machine learning models identifying AD.

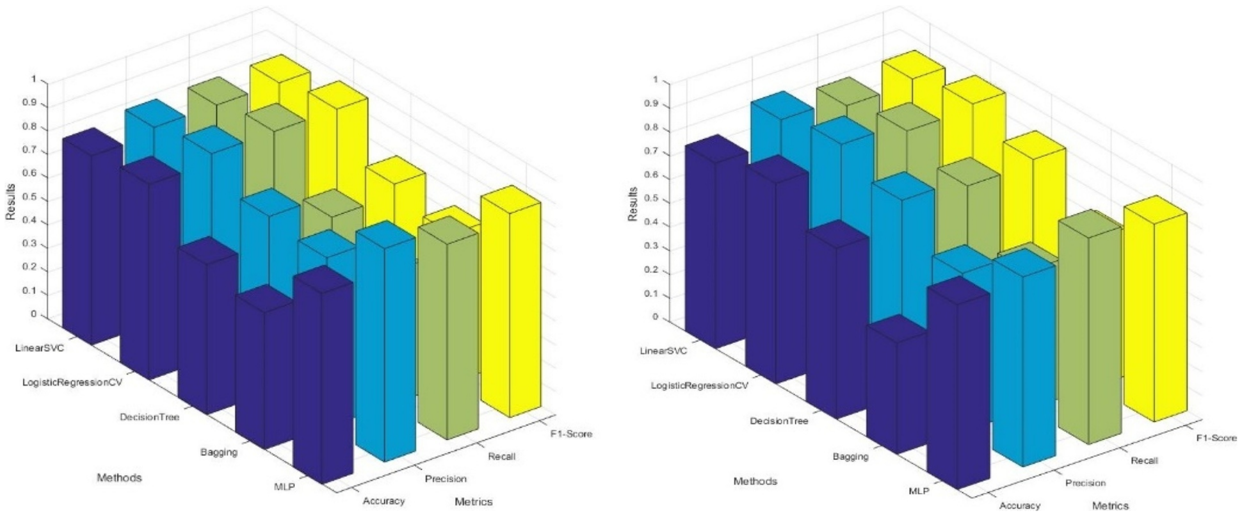

Fig. 8. Performance results of different methods on the VBSD dataset (left) and the Dem@Care dataset (right).

Table 6
Experimental results of different methods based on VBSD dataset.

Methods	Accuracy	Precision	Recall	F1-Score
LinearSVC	0.806	0.833	0.833	0.833
LogisticRegressionCV	**0.833**	**0.869**	**0.869**	**0.869**
DecisionTree	0.639	0.750	0.652	0.698
Bagging	0.583	0.722	0.565	0.634
MLP	0.813	0.909	0.833	0.869

training sets and testing sets are in constant change, and thus DecisionTree classifier performs poorly in this simulation experiment. The base classifier of Bagging chooses DecisionTree classifier, and the performance of Bagging mainly depends on the stability of the base classifier. As the result, the classification performance of Bagging is also poor. The differences in language and pronunciation between countries are not considered in combining the two datasets in simulation experiment.

Table 7
Experimental results of different methods based on Dem@Care dataset.

Methods	Accuracy	Precision	Recall	F1-Score
LinearSVC	0.781	0.869	0.833	0.851
LogisticRegressionCV	**0.844**	**0.913**	**0.875**	**0.894**
DecisionTree	0.719	0.826	0.792	0.809
Bagging	0.469	0.667	0.583	0.622
MLP	0.778	0.800	0.869	0.833

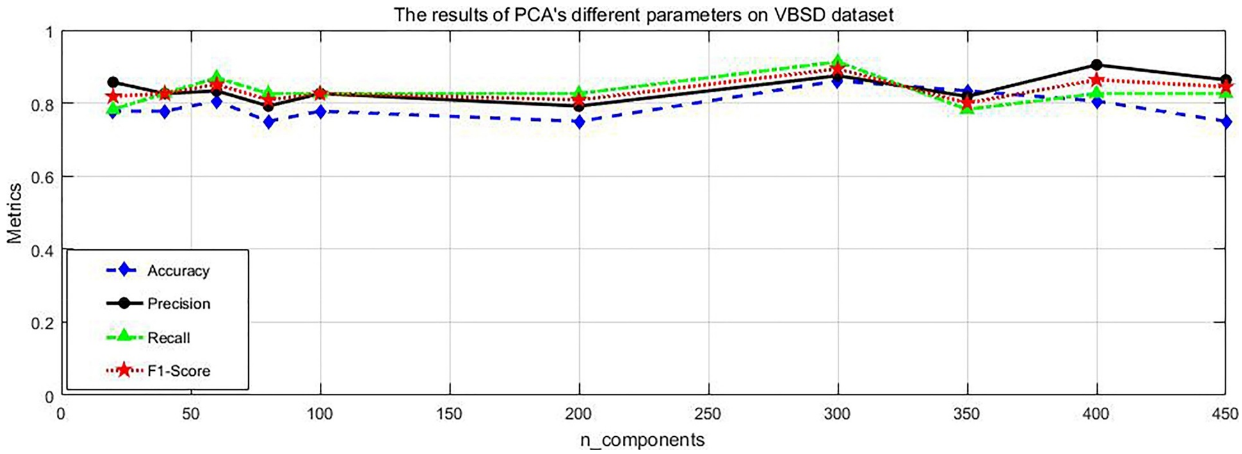

Fig. 9. The influence of different parameters on the metrics.

Through the above simulation experiments, the best performing method is the LogisticRegressionCV method. In order to further optimize this method, different parameter values for PCA dimension reduction experiments are carried out to achieve higher identification performance. Fig. 9 shows the performance metrics with different dimensions of the respective 450-dim reduced dimension 20-dim. As it can be seen that when n_component = 300 (the number of feature dimensions), accuracy achieves the best identification accuracy, the results are an accuracy of 86.1%, precision of 87.5%, recall of 91.3%, and f1-score of 89.4%, respectively. The identification results are different when the dimensions are different. Therefore, when conducting experiments, the parameter adjustment is needed, which can optimize the method to get better results.

4. Conclusions

This paper proposed a new method that used the spectrogram features extracted from speech data to identify AD, which could help families understand the disease development of patients in an earlier stage, in order to take measures in advance to delay the disease development. In addition to the proposed new method, a new dataset (VBSD dataset) was collected. Most of the existing work was to extract the numerical parameters of a person's speech, including speech rate, number of pauses, reaction time, etc., which were more limited by the size of dataset. The method proposed in this paper divided a person's speech data into segments and added dataset samples. In the process of experimenting with the speech data of AD and HC, different machine learning methods were used for experiments. The final experimental results showed that the LogisticRegressionCV model had achieved the best performance. In the future work, a speech system APP can be developed to assist families to observe the status of the elderly. More data can be collected to conduct experiments to improve the accuracy of identifying each category. The identification category can be further divided into different stages of AD for experiment.

Acknowledgments

The authors wish to express the gratitude for the Dem@Care FP7 project for making audio dataset and all of subjects available, which facilitated our experiments. Project supported by the Natural Science Research of the Anhui Higher Education Institution of China (Grant No. KJ2019ZD44, KJ2019A0647).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.simpat.2019.102023.

References

[1] C. Patterson, World Alzheimer Report 2018. 'An Analysis of Prevalence, Incidence, Cost and Trends', The State of the Art of Dementia Research: New Frontiers, Alzheimer's Disease International, 2018.
[2] Shuya Wang, Kai Liu, Hong Liu, Baoqing Gui, Depression among caregivers of patients with alzheimer's disease, World Latest Med. Inf. 17 (2) (2017) 213.
[3] Alzheimer Alois, Über eine eigenartige erkrankung der hirnrinde, Allgemeine Zeitschrift für Psychiatrie und Psychisch-Gerichtlich Medizin 64 (1) (1907) 146–148.
[4] Donghuan Lu, Karteek Popuri, Gavin Weiguang Ding, Rakesh Balachandar, Mirzal Faisal Begand Alzheimer's Disease Neuroimaging Initiative, Multimodal and multiscale deep neural networks for the early diagnosis of Alzheimer's disease using structural MR and FDG-PET images, Sci. Rep. 8 (1) (2018) 1–13.
[5] J. Shi, X. Zheng, Y. Li, Q Zhang, S. Ying, Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer's disease, IEEE J. Biomed. Health Inf. 15 (1) (2018) 67–79.
[6] A. Ortiz, F. Lozano, J.M. Gorriz, J. Ramirez, F.J Martinez Murciaand Alzheimer's Disease Neuroimaging Initiative, Discriminative sparse features for Alzheimer's disease diagnosis using multimodal image data, Curr. Alzheimer's Res. 15 (1) (2018) 67–79.
[7] Q. Li, X. Wu, F. Xie, K. Chen, L. Yao, J. Zhang, X. Guo, R Liand Alzheimer's Disease Neuroimaging Initiative, Aberrant connectivity in mild cognitive impairment and Alzheimer disease revealed by multimodal neuroimaging data, Neurodegener. Dis. 18 (1) (2018) 5–18.
[8] E.E. Tripoliti, D.I. Fotiadis, M. Argyropoulou, A supervised method to assist the diagnosis and classification of the status of Alzheimer's disease using data from an fMRI experiment, Conference: International Conference of the IEEE Engineering in Medicine & Biology Society IEEE Engineering in Medicine & Biology Society Conference, Conf Proc IEEE Eng Med Biol Soc, 2008.
[9] S. Sarraf, G... Tofighi, Deep learning-based pipeline to recognize Alzheimer's disease using fMRI data, Future Technologies Conference 2016, IEEE, 2016.
[10] X.W. Gao, R... Hui, A deep learning based approach to classification of CT brain images, 2016 SAI Computing Conference (SAI), IEEE, 2016.
[11] F.C. Morabito, M. Campolo, C. Ieracitano, et al., Deep convolutional neural networks for classification of mild cognitive impaired and Alzheimer's disease patients from scalp EEG recordings, International Forum on Research and Technologies for Society and Industry Leveraging A Better Tomorrow, IEEE, 2016, pp. 1–6.
[12] Mohsin Raza, M. Awais, W. Ellahi, N. Aslam, Huan X. Nguyen, H. Le-Minh, Diagnosis and monitoring of Alzheimer's patients using classical and deep learning techniques, Expert Syst. Appl. 136 (2019) 353–364.
[13] Simeon Spasov, Luca Passamonti, Andrea Duggento, Pietro Liò, Nicola Toschi, A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease, Neuroimage 189 (2019) 276–287.
[14] Greta Szatloczki, Ildiko Hoffmann, Veronika Vincze, Janos Kalman, Magdolna Pakaski, Speaking in Alzheimer's disease, is that an early sign? Importance of changes in language abilities in Alzheimer's disease, Front Aging Neurosci. 7 (2015) Oct 20.
[15] K.A. Bayles, D.R. Boone, The potential of language tasks for identifying senile dementia, J. Speech Hear Disord. 47 (2) (1982) 210–217.
[16] K.C. Fraser, J.A. Meltzer, F Rudzicz, Linguistic features identify Alzheimer's disease in narrative speech, J. Alzheimer's Dis. 49 (2016) 407–422.
[17] V. Rentoumi, L. Raoufian, S. Ahmed, C.A. de Jager, P Garrard, Features and machine learning classification of connected speech samples from patients with autopsy proven Alzheimer's disease with and without additional vascular pathology, J. Alzheimer's Dis. 42 (2014) 3–17.
[18] Colin G. Harrison, Peter R. Williams, A systems approach to natural disaster resilience, Simul. Modell. Pract. Theory 65 (2016) 11–31.
[19] Ondřej Háva, Miroslav Skrbek, Pavel Kordík, Supervised two-step feature extraction for structured representation of text data, Simul. Modell. Pract. Theory 33 (2013) 132–143.
[20] Tarik Al-Ani, Yskandar Hamam, Redouane Fodil, Frédéric Lofaso, Daniel Isabey, Using hidden Markov models for sleep disordered breathing identification, Simul. Modell. Pract. Theory 12 (2004) 117–128.
[21] Ahmad B.A. Hassanat, Mouhammd Alkasassbeh, Mouhammd Al-awadi, Esra'a A.A. Alhasanat, Color-based object segmentation method using artificial neural network, Simul. Modell. Pract. Theory 64 (2016) 3–17.
[22] Vikram Venkatadri, Venkatesh Arasanipalai Raghavan, Varun Kesavakumaran, Sarah S. Lam, Krishnaswami Srihari, Simulation based alternatives for overall process improvement at the cardiac catheterization lab, Simul. Modell. Pract. Theory 19 (2011) 1544–1557.
[23] Yinong Chen, Hualiang Hu, Internet of intelligent things and robot as a service, Simul. Modell. Pract. Theory 34 (2013) 159–171.
[24] Yinong Chen, Service-Oriented Computing and System Integration: Software, IoT, Big Data, and AI As Services, 6th edition, Kendall Hunt Publishing, 2018.
[25] A. Karakostas, A. Briassouli, K. Avgerinakis, I. Kompatsiaris, M. Tsolaki, The dem@care experiments and datasets: a technical report, (2016) FP7-288199 Dem@Care.
[26] A. Satt, A. Sorin, O. Toledoronen, O. Barkan, I. Kompatsiaris, A. Kokonozi, et al., Evaluation of speech-based protocol for detection of early-stage dementia, Interspeech (2013).
[27] Alex SorinIBM Research -Haifa, Voice Analytics for Dementia Assessment, DemAAL, Chania, Greece, 2013 September 2013.
[28] A. K?Nig, A. Satt, A. Sorin, R. Hoory, O. Toledo-Ronen, A. Derreumaux, et al., Automatic speech analysis for the assessment of patients with predementia and alzheimer\"s disease, Alzheimer\"s Dementia 1 (1) (2015) 112–124.