

Article

Deep Learning of Speech Data for Early Detection of Alzheimer's Disease in the Elderly

Kichan Ahn ^{1,†} , Minwoo Cho ^{2,3,4,†}, Suk Wha Kim ^{3,5}, Kyu Eun Lee ^{3,6} , Yoojin Song ⁷, Seok Yoo ⁸, So Yeon Jeon ⁹ , Jeong Lan Kim ^{9,10}, Dae Hyun Yoon ¹¹ and Hyoun-Joong Kong ^{2,3,4,*} 

- ¹ Interdisciplinary Program in Medical Informatics Major, Seoul National University College of Medicine, Seoul 03080, Republic of Korea; mikeahn@snu.ac.kr
 - ² Department of Transdisciplinary Medicine, Seoul National University Hospital, Seoul 03080, Republic of Korea; chovis@snuh.org
 - ³ Medical Big Data Research Center, Seoul National University College of Medicine, Seoul 03080, Republic of Korea; kimsww@snu.ac.kr (S.W.K.); kyueunlee@snu.ac.kr (K.E.L.)
 - ⁴ Department of Medicine, Seoul National University College of Medicine, Seoul 03080, Republic of Korea
 - ⁵ Department of Plastic Surgery and Institute of Aesthetic Medicine, CHA Bundang Medical Center, CHA University, Seongnam 13496, Republic of Korea
 - ⁶ Department of Surgery, Seoul National University Hospital and College of Medicine, Seoul 03080, Republic of Korea
 - ⁷ Department of Psychiatry, Kangwon National University, Chuncheon 24289, Republic of Korea; o2pamu@naver.com
 - ⁸ Unidocs Inc., Seoul 03080, Republic of Korea; tobewise@unidocs.co.kr
 - ⁹ Department of Psychiatry, Chungnam National University Hospital, Daejeon 30530, Republic of Korea; iris870422@gmail.com (S.Y.J.); thomasign@gmail.com (J.L.K.)
 - ¹⁰ Department of Psychiatry, Chungnam National University College of Medicine, Daejeon 30530, Republic of Korea
 - ¹¹ Department of Psychiatry, Healthcare System Gangnam Center, Seoul National University Hospital, Seoul 03080, Republic of Korea; gong@snuh.org
- * Correspondence: gongcop7@snu.ac.kr; Tel.: +82-2-2072-4492
† These authors contributed equally to this work.



Citation: Ahn, K.; Cho, M.; Kim, S.W.; Lee, K.E.; Song, Y.; Yoo, S.; Jeon, S.Y.; Kim, J.L.; Yoon, D.H.; Kong, H.-J. Deep Learning of Speech Data for Early Detection of Alzheimer's Disease in the Elderly. *Bioengineering* **2023**, *10*, 1093. <https://doi.org/10.3390/bioengineering10091093>

Academic Editors: Crescenzo Gallo, Gianluca Zaza and Richard Bayford

Received: 4 June 2023

Revised: 16 August 2023

Accepted: 24 August 2023

Published: 18 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Background: Alzheimer's disease (AD) is the most common form of dementia, which makes the lives of patients and their families difficult for various reasons. Therefore, early detection of AD is crucial to alleviating the symptoms through medication and treatment. Objective: Given that AD strongly induces language disorders, this study aims to detect AD rapidly by analyzing the language characteristics. Materials and Methods: The mini-mental state examination for dementia screening (MMSE-DS), which is most commonly used in South Korean public health centers, is used to obtain negative answers based on the questionnaire. Among the acquired voices, significant questionnaires and answers are selected and converted into mel-frequency cepstral coefficient (MFCC)-based spectrogram images. After accumulating the significant answers, validated data augmentation was achieved using the Densenet121 model. Five deep learning models, Inception v3, VGG19, Xception, Resnet50, and Densenet121, were used to train and confirm the results. Results: Considering the amount of data, the results of the five-fold cross-validation are more significant than those of the hold-out method. Densenet121 exhibits a sensitivity of 0.9550, a specificity of 0.8333, and an accuracy of 0.9000 in a five-fold cross-validation to separate AD patients from the control group. Conclusions: The potential for remote health care can be increased by simplifying the AD screening process. Furthermore, by facilitating remote health care, the proposed method can enhance the accessibility of AD screening and increase the rate of early AD detection.

Keywords: Alzheimer's disease; mental status and dementia tests; early diagnosis; speech acoustics; deep learning; digital healthcare

1. Introduction

Alzheimer's disease (AD), the most common form of dementia, is a neurodegenerative disease characterized by cognitive decline [1–3]. Among the various causes of dementia, AD accounts for 60–70% of all cases [4]. The increase in life expectancy has been associated with a steady increase in the population with AD [5,6]. After the age of 65, the likelihood that a person will develop AD doubles every five years. Therefore, the number of patients with dementia will reportedly be over three times higher by 2050 compared to 2010 [7,8].

From the early stages to the most severe cases, the symptoms of AD include decreased spatial awareness, lack of concentration, and distraction [9,10]. Other major symptoms include memory impairment and deficits in language skills [11–13]. Additionally, physical function decreases, which makes it difficult to perform daily activities; therefore, patients lose autonomy and become dependent on others for care [13,14]. However, deterioration in language ability can impair communication, thereby making daily life difficult for AD patients and caregivers [14,15]. Over time, family members of AD patients may experience increased physical and emotional exhaustion [16,17]. Moreover, owing to the high costs of all care, diagnosis, and pharmacological treatment, AD is one of the most expensive chronic diseases [15,18]. For example, the cost of caring for patients with AD and other forms of dementia is over twice that of patients of the same age suffering from cancer and 74% higher than those with cardiovascular diseases [19,20]. Consequently, patients and their families incur a financial burden [14,16–18].

Currently, there is no cure for AD, and it is considered a very serious disease [2,4]. However, if detected early, the progression of symptoms can be delayed or alleviated with medication [21,22]. A definitive diagnosis of AD includes diagnostic techniques such as genetic tests, cerebrospinal fluid tests, positron emission tomography (PET), and magnetic resonance imaging (MRI), which can be costly and invasive [23,24]. Therefore, they are unsuitable for early diagnosis [18,23,24]. In addition, various standards for AD diagnosis exist, but most depend on the results of tests performed by experts [25,26]. Furthermore, AD has social consequences, such as the cost to the national economy. Therefore, there is increasing interest in the development of simple screening techniques that can provide an easy and convenient diagnosis that is accessible and low-cost [27–29].

One possible solution is using speech analysis and processing to detect changes in language ability, which can facilitate the early detection of AD [30,31]. These changes could be a key indicator for the preclinical stages of AD [12,30] and for patients experiencing greater difficulty speaking as the disease progresses [13]. AD patients may show speech-related symptoms such as hesitation, frequent pauses, blurred pronunciation, tremors, light stuttering, the use of irregular words, reduced verbal fluency, changes in the rhythm of speech, deviation from simple grammatical and lexical rules, slow or irregular breathing, and an inability to control breathing [12,31]. Moreover, there is a close relationship between language ability and cognitive ability [32,33]. These characteristics can be used as an initial indicator to distinguish between AD-related anomia and non-AD pathology [12,30]. In terms of the order of symptom manifestation, it can be considered that language impairment occurs before memory impairment; therefore, it can be a good predictor of early AD [32,33].

Although changes in acoustic and vocal rhythm may be imperceptible to the human ear, advances in automatic speech analysis technology have made it possible to identify and effectively extract these acoustic and temporal parameters [14,34]. Speech biometrics or automatic speech analysis are considered ideal tools for assessing cognitive deficits or changes in older adults, as these methods are capable of recording speech planning, sequencing, and performance in real time [35]. Recently, various studies have attempted to classify spoken language using different speech processing techniques and algorithms to identify the early signs of cognitive decline [14,27,35,36].

In addition to the symptoms of language disorders, the vocabulary level, complexity of syntactic structure, and use of irregular words are significantly affected by factors such

as age, educational background, and cognitive ability; therefore, it is difficult to use these predictors as indicators of early AD [37]. In contrast, the frequency of hesitation, impaired affective prosody, emphasis of specific syllables, changes in tempo or timing, differences in pitch and intonation, and irregular breathing can be used as indicators in speech analysis and processing of voice signals [38–40]. Language analysis is important owing to its suitability for classification; some studies have shown that it can be used to distinguish between people with and without AD with over 91.2% accuracy [41,42].

The COVID-19 pandemic has increased the demand for remote diagnosis and management of AD [43,44]. The existing mode of mini-mental state examination for dementia screening (MMSE-DS) requires patients to visit medical institutions for in-person screenings. Therefore, the restrictions imposed due to the COVID-19 pandemic have made early AD diagnosis difficult [45,46]. To satisfy the changing requirements, government agencies are planning to introduce a remote AD management system that will enable elderly people with reduced mobility to undergo dementia screening examinations at home [47–49]. Hence, there is a pressing need to develop measures with greater accuracy and efficiency for remote AD diagnostic screening [48,50,51]. Previously, remote dementia management systems were implemented over the phone [29,32,35]. Also, in recent years, there has been active research on the delivery of telemedicine via smart devices and applications [45,46]. Various smart devices are highly suitable for the diagnosis and remote management of AD, considering they can quickly and easily capture voice and image data and, in some cases, record basic bio-signals [46,49,52].

In this study, MMSE-DS is performed with AD patients and healthy adults to select factors that are significant for classifying dementia patients based on case records. After pre-processing the voice data obtained for each item, mel-frequency cepstral coefficients (MFCCs) are used to produce a spectrogram by arranging the coefficients in a specific order defined by the authors and synthesizing them into a single image. These images are used for training and are applied to different deep learning models to obtain high accuracy. Furthermore, we establish the criteria for the selection of factors suitable for analysis based on the MMSE-DS and voice data. Lastly, we verify that the proposed method can be used to diagnose AD with high accuracy compared to MMSE-DS by utilizing voice signals that can be easily acquired by exchanging simple questions and answers online using deep learning methods without establishing a special examination system for AD screening.

2. Methods

2.1. Patient Information

The voice data of AD patients (experimental group) and healthy adults (control group) were obtained by applying MMSE-DS to 88 adults aged 50–75 years who expressly indicated their voluntary intention to participate (Table 1). The study was conducted at Chungnam National University Hospital between 1 April 2019 and 23 December 2019. The experimental group included those who had been diagnosed with AD in the last three months and those who had a clinical dementia rating between 0.5 and 2.

Table 1. Status of age and gender distribution related to research data.

Age	AD Patients		Healthy Adults	
	Male	Female	Male	Female
50–59	0	0	1	8
60–69	6	12	7	12
70–75	2	20	4	8
Total	8	32	12	28
	40		40	

All experimental groups were confirmed to be AD patients through MRI, blood tests, and neuropsychological tests. Patients with a rating > 3 or those who were unable to un-

dergo screening were excluded from this study. The control group was recruited via a participant recruitment notice, and those who were deemed capable of conducting the screening test were identified through a simple interview with a specialist. Finally, 42 patients who scored at least 26 points were selected as the control group.

This study was approved by the institutional review board of Chungnam National University Hospital (human subject study, prospective study, observational study, controlled study) (IRB approval number: CNUH2019-02-068), and the study protocol adhered to the ethical guidelines of the 1975 Declaration of Helsinki. The MMSE-DS for all participants was performed by one trained clinical psychologist. The entire MMSE-DS process for each participant was recorded as a video and saved as an MP4 file. Out of the 88 participants recruited, eight were excluded because they expressed their wish to withdraw. Of the remaining 80 participants, 24 were men (11 in the experimental group) and 56 were women (29 in the experimental group), i.e., accounting for 30 and 70% of the total participants, respectively. The mean age of the participants was 68.8 years. The average length of education was 9.36 years; 11.2 years for men and 8.27 years for women.

2.2. Clinical Data Collection

The screening test was conducted using MMSE-DS, the most commonly used AD screening method by public health centers in South Korea. MMSE-DS includes the following elements: temporal orientation, spatial orientation, memory registration, attention, recall, visual denomination (naming), following a three-stage command, phrase repetition, visuospatial construction, reading, writing, comprehension, and judgment, and analysis of scores [53–55]. Considering this study only used voice characteristics, the results for the questions that required spoken answers were used [56,57]. Table 2 summarizes the MMSE-DS composition and questions.

Considering this study aims to perform a classification of AD patients by analyzing voice data, we assumed that it would be advantageous and efficient to exclude some of the results instead of using them all [33,35,36,57]. The selection of the questionnaires was based on the expert advice of a focus group interview conducted by psychiatrists at Chungnam National University Hospital. The questionnaires recommended excluding some questions on temporal orientation, the ability to follow a command, spatial orientation, and visuospatial construction as they required judgment, visual denomination, and abstract thinking. In particular, questions on the ability to follow a command and visuospatial construction did not require a verbal response; hence, they were not suitable for this study [32,58].

Next, based on the experts' advice, items such as spatial orientation, temporal orientation, judgment, and visual denomination were gradually excluded to see the results. Finally, twelve questions were determined to be most suitable for distinguishing AD patients from the control group, which are marked with an asterisk in Table 2. They include three items for memory registration, five items for attention and calculation, one item for delayed recall, and one item for temporal orientation [53,54,59,60]. Although the responses to all 28 questions were obtained, only twelve results were used for the subsequent MFCC deep learning and spectrogram.

MMSE-DS was performed once per participant, and the voice responses were acquired for 88 participants. All the recorded data were valid. After screening, eight patients expressed their intention to withdraw from the study, and their results were excluded. Once the consent of the remainder of participants was obtained, the video and audio data were recorded for the remaining 80 participants using a webcam (Logitech BRIO 4K webcam) with supported 30 fps, ultra-high definition, and 4096×2160 resolution.

Among the recorded voice responses, the parts corresponding to the participants' answers were edited and saved as a wave file using the Cubase audio editor.

Table 2. Questionnaires for the MMSE-DS used in this study.

Number	Items	Individual Questions
1-1	Temporal orientation	What year are we in now?
1-2		What is the season?
1-3		What is the date today?
1-4		What day of the week is it?
1-5 *		What month are we in now?
2-1	Spatial orientation	What city are we in?
2-2		What borough are we in?
2-3		What 'dong' (one of the administrative divisions) are we in?
2-4		What floor of the building are we on?
6		What is the name of this place?
10	Following a three-stage command	Please follow what I say and as it will be told only once, please listen carefully and follow accordingly. I will give you a piece of paper. Please take this piece of paper in your right hand, fold it in half with both hands, and place it on your lap.
11 *	Memory registration	I am going to name three objects. After I have said them, I want you to repeat them. Please remember what they are because I will ask you to name them again in a few minutes: tree (11-1 *), car (11-2 *), hat (11-3 *). Could you name the three items you have just heard?
12-1 *	Attention and calculation	What is one hundred minus seven?
12-2 *		Yes. Then, what is the result after subtracting seven from the value?
12-3 *		Yes. Then, what is the result after subtracting seven from the value?
12-4 *		Yes. Then, what is the result after subtracting seven from the value?
12-5 *		Yes. Then, what is the result after subtracting seven from the value?
13 *	Delayed recall	What are the three objects I asked you to remember a few moments ago? Tree (13-1 *), car (13-2 *), hat (13-3 *).
14-1	Visual denomination	(Showing a watch) What is this called?
14-2		(Showing a pencil) What is this called?
15	Phrase repetition	Please listen carefully to what I say and repeat accordingly. Please note that only one attempt will be allowed. Please listen carefully and repeat after I finish. Ganjang Gongjang Gongjangjang (Translation: head of the soy source factory, used for checking pronunciation)
16	Visuospatial construction (Copying interlocking pentagons)	Please see the interlocking pentagons here and copy the drawing in the following blank section.
18	Judgment	Why do you need to wash your clothes?
19		Could you explain what "many a mickle makes a muckle" means?

* Questions selected for MFCC deep learning and spectrogram.

2.3. Data Preprocessing

The audio data corresponding to the responses to the twelve selected questions was extracted from the MMSE-DS videos using the ffmpeg tool with a sampling rate of 44.1 kHz and a stereo channel format. Then, the MP4 file format was used to minimize information loss in the audio. The audio response information was set to 3 s as the participants' answers were completed within 1–2 s in most cases. When the response was completed in less than 3 s, silence was added to create a file with a total length of 3 s. Participants in the AD group were often unable to answer the questions. In this case, they were given 30 s; if there was still no answer, the next question was asked. If a participant could not answer the

questions, the entire 30 s period was treated as silence. In these cases, when the voice data was converted to a spectrogram, the length of the dataset became long without containing significant meaning. Therefore, to construct efficient training datasets, a 3 s clip of the 30 s silence was extracted and converted to a spectrogram. As all participants responded within 1–2 s, “no response”, which was expressed as 3 s of silence, was sufficiently distinguishable. Features from one wave file were converted to one image using the MFCC, and twelve MFCC images were extracted from twelve wave files. Next, these images were combined and reconstructed to form one image file. Figure 1 shows a conceptual diagram of the pre-processing procedure.

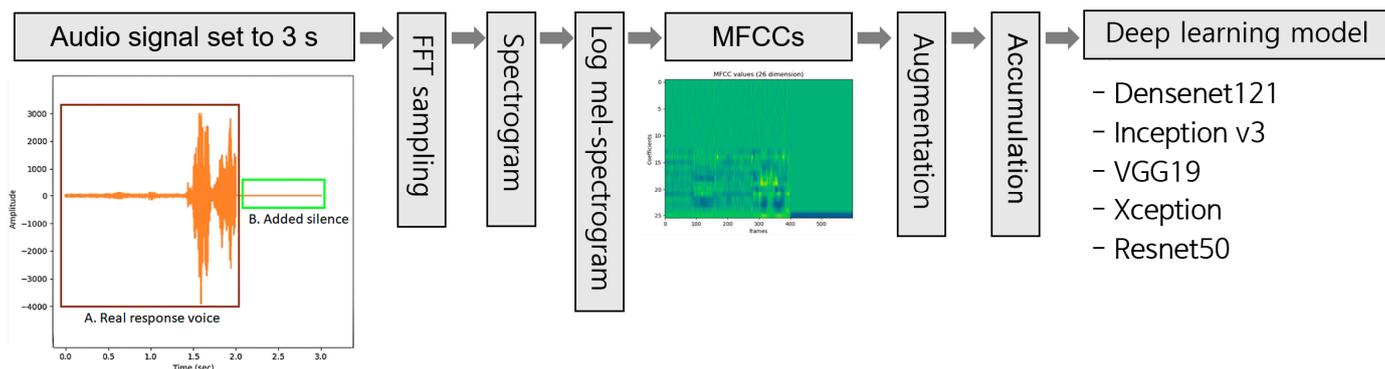


Figure 1. Overall process flow of the study.

2.4. Spectrogram

Extracting features from a wave file using MFCCs is a common technique used when processing voice signals [61,62]. The sampling rate of the wave file was 44.1 kHz, which indicated that there were 44,100 signals s^{-1} . Each wave file had a duration of 3 s, which gives 132,300 signals per file. If 0.025 s of audio information is defined as one frame of the wave file, then there are 1103 signals per frame (rounded to the nearest whole number). The period of frame extraction was 0.01 s; hence, the frame information was extracted by skipping $0.01 \times 44,100 = 441$ signals. Therefore, if MFCC information was extracted from 132,300 signals for 3 s, a width of 599 MFCC images was obtained. The MFCC image was 26 pixels high because 13 MFCC feature values and 13 MFCC first derivative values were extracted from each frame. Therefore, the MFCC image generated for each 3 s response wave file was 599×26 pixels. By compiling the results for all twelve questions on the vertical axis, the learning data (MFCC) comprising 599×312 pixels were generated for each participant’s response. This description is shown in Figure 2.

2.5. Deep Learning Algorithms

Figure 1 shows the overall process, wherein the MFCC images are generated from the audio data and used to train and test deep learning models. To improve the accuracy during training, ten-fold data augmentation was performed by translating the data horizontally and changing the brightness [63]. Because speech characteristics are expressed as patterns and textures in MFCC images, augmentation methods that induce morphological transformation were avoided. Further, we experimentally examined the range in which the change in brightness value in MFCC image augmentation is suitable for learning in deep learning. The brightness was changed by 80–120% compared to that of the original image, in 5% increments. The “horizontal shift” led to a delayed response effect, and the changes according to the shift range were validated and applied so that the effect was not excessive. The horizontal shift was increased to $\pm 25\%$ in 5% increments.

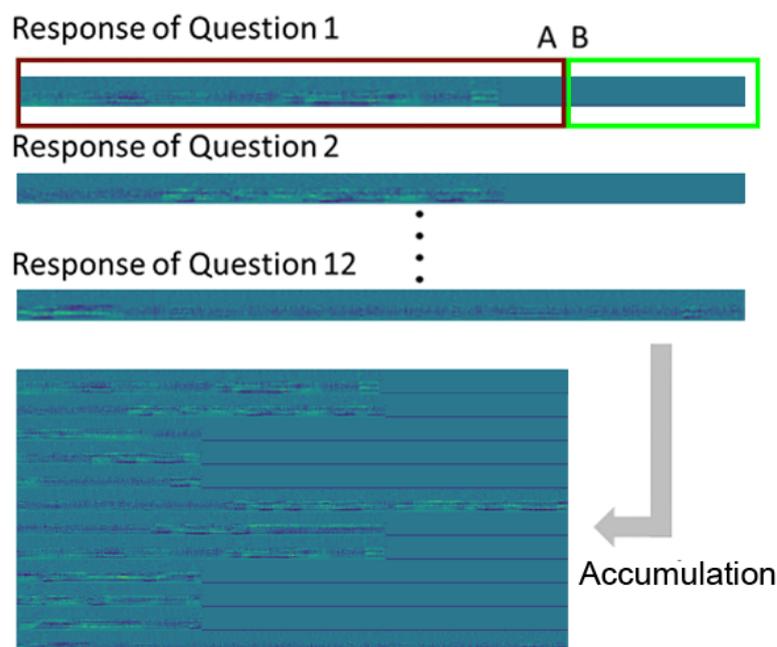


Figure 2. MFCC image generation of each participant's response. (A) The answers to each question were standardized to a duration of 3 s. (B) If there was no answer at the end of this period, silence was added to the entire duration.

In order to perform accurate augmentation, we confirmed the section where the Densenet121 model showed the highest performance after augmenting data for brightness and horizontal shift, respectively. As the convolutional network incorporates shorter connections between layers closer to the input and those closer to the output, the network becomes deeper, more accurate, and capable of learning more efficiently. Densenet121 connects each layer to every other layer in a feed-forward manner. In traditional convolutional networks, each layer has L connections (one connection to the next layer), whereas Densenet121 has $L \times (L + 1)/2$ connections. For each layer, all previous feature maps are used as inputs, and its own feature map is utilized as inputs for all subsequent layers. Therefore, the advantages of Densenet121 include: (1) alleviating the vanishing-gradient problem and enhancing feature propagation. (2) encouraging feature reuse and reducing the number of parameters. Considering these characteristics, the Densenet121 model, which is suited for MFCC analysis, was selected in this experiment [64,65]. Through this process, we performed data augmentation techniques that were advantageous for deep learning training. The corresponding results are presented in Appendix A (Tables A1 and A2). Based on these results, we finally applied brightness $\pm 15\%$ and right shift 20% augmentation to our training dataset. The eighty MFCC images obtained from the participants were augmented tenfold; hence, the results comprised 400 images each for healthy adults and AD patients.

To classify healthy adults and AD patients using MFCC images, training and predictions were performed using five deep learning models, which are representative CNN algorithms for image classification and include Densenet121, Inception v3, VGG19, Xception, and Resnet50 [66]. In the MFCC image, the signal characteristics were expressed in the form of a pattern or texture. Therefore, deep learning models suitable for image pattern or texture type classification were selected [67–69]. The performance of the five CNN algorithms was evaluated using the five-fold cross-validation and hold-out methods, wherein the datasets were divided in a ratio of 8:1:1 for training, validation, and testing, respectively [70].

The training for each algorithm was stopped early to avoid overfitting by checking the training and validation losses three times. AdaMax was used as an optimizer to optimize the deep learning models, with a learning rate of 1×10^{-6} and a batch size of 4. For all five types of models, the number of epochs was equally applied to 60, the number of training

data was 640, and the length of the test data was 16. In the case of five-fold cross-validation, all models were iterated five times and the validation splits were set at 0.1.

3. Results

Performance Comparison

In order to acquire accurate results for the five deep learning algorithms, the performance results of the hold-out and cross-validation methods were obtained. In this experiment, we focused on metrics such as sensitivity, which is the value predicted by the model as AD patients among actual AD patients, and positive predictive value (PPV), which represents the ratio of actual AD patients among predicted AD patients [71]. Additionally, the F1-score was utilized to assess the model's balanced performance between positive and negative predictions. These metrics provide valuable insights into the performance of the classification model in terms of correctly identifying positive instances, the accuracy of positive predictions, and the overall balance between precision and recall. Furthermore, metrics such as specificity and negative predictive value (NPV), which is a value representing the actual proportion of normal people among normal people, were calculated and shown in Tables 3 and 4. The confusion matrix for each algorithm is in Figures A1 and A2 of Appendix A.

Table 3. Performance metrics of the five deep learning models for the five-fold cross-validation.

Model Name	Sensitivity	Specificity	Accuracy	PPV	NPV	F1-Score	AUC
Densenet121	0.9550	0.8333	0.9000	0.8791	0.9314	0.9139	0.9243
Inception v3	0.9305	0.8099	0.8750	0.8556	0.9179	0.8887	0.9177
VGG19	0.9750	0.8236	0.9000	0.8494	0.9778	0.9013	0.8886
Xception	0.9455	0.3183	0.6000	0.5855	0.9143	0.6997	0.8349
Resnet50	0.8944	0.8979	0.9000	0.8994	0.9042	0.8956	0.9286

Table 4. Performance metrics of the five deep learning models for the hold-out validation.

Model Name	Sensitivity	Specificity	Accuracy	PPV	NPV	F1-Score	AUC
Densenet121	1.0000	0.7143	0.8750	0.8182	1.0000	0.9000	0.9048
Inception v3	0.9000	0.8333	0.8750	0.9000	0.8333	0.9000	0.9500
VGG19	1.0000	0.6250	0.8125	0.7273	1.0000	0.8421	0.9219
Xception	1.0000	0.2500	0.6250	0.5714	1.0000	0.7273	0.8594
Resnet50	0.8889	1.0000	0.9375	1.0000	0.8750	0.9412	0.9524

As a result, in the five-fold cross-validation, Densenet121 showed the highest performance with a sensitivity of 0.9550, an accuracy of 0.9000, a PPV of 0.8791, an F1-score of 0.9139, and an AUC of 0.9243. As shown in the AUC graph (Figure 3), Resnet50 and Inception v3 also showed high performance. Likewise, in the results of the hold-out validation, Densenet121, Inception v3, and Resnet50 showed high performance in terms of sensitivity, accuracy, PPV, F1-score, and AUC. Figure 3 compares the results obtained using the hold-out and five-fold cross-validation methods. The gray dashed curve represents the AUC of MMSE-DS [59].

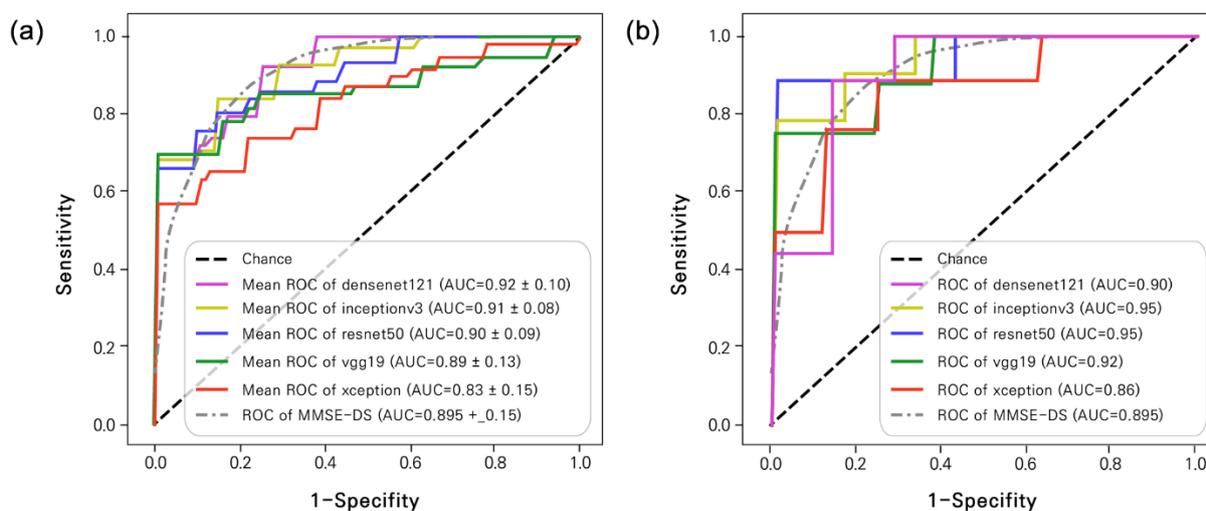


Figure 3. Receiver operating characteristic (ROC) results. (a) Five-fold cross-validation and (b) Hold-out methods. The gray line indicates the AUC of MMSE-DS [60].

4. Discussion

This study presented a method of screening AD patients using only voice data. Voice data were collected, pre-processed, normalized, and converted into a spectrogram to obtain MFCC images. Deep learning models are commonly trained with the MFCC transition to utilize the advantages of sound signals and non-verbal elements [71–74]. MFCC features have been widely used in voice classification tasks, as they have been shown to perform well in terms of robustness and discrimination power. These features are based on the mel-frequency scale, which is a non-linear frequency scale that is closely related to the non-verbal elements. For this reason, we used MFCC and two-dimensional spectrogram images [75–77]. The MMSE-DS assessment tool, the most widely used method for interviewing AD patients in public health centers in South Korea, was used to acquire the voice data. The MMSE-DS comprised 28 questions that were used to screen patients for their attention and calculation, temporal orientation, memory registration, and delayed recall.

Responses were obtained for all questions with MMSE-DS. At the beginning of the experiment, we trained the deep learning model using all the questions. We attempted to train the model in various ways, but the ROC value was very low at 0.6. Therefore, we judged that it is more effective to select valid questions and use them for model training than to use all questionnaires. For this reason, not all of these responses were converted into MFCC images; only 12 questions were finally selected based on expert advice from a focus group interview conducted by psychiatrists at Chungnam National University Hospital. The excluded questions were based on temporal orientation, ability to follow a command, spatial orientation, and visuospatial construction, as they required judgment, visual denomination, and abstract thinking. In particular, the ability to follow a command and visuospatial construction questions did not require a verbal response, and hence, they were not suitable for this study [32,58].

The answers to each question were standardized to a duration of 3 s. One MFCC image was generated, and the answers to all 12 questions were compiled to generate a training image. The responses were standardized to 3 s because none of the responses in either group exceeded this duration. However, the AD patients were unable to answer some questions, in which case a waiting time of 30 s was provided. If there was no answer at the end of this period, silence was added to the entire duration. The texture of the MFCC image of the part processed as silence was different from the voice signals, and if all 30 s of the wait time were used, the ratio of the texture of silence would become unnecessarily high. Considering that these factors hinder learning performance, the voice data were standardized to 3 s, which was found to be an appropriate time to receive an answer.

In this study, the AD classification accuracy of the deep learning model was measured using the five-fold cross-validation and hold-out validation methods. In the five-fold cross-validation method, the Densenet121 model exhibited the highest overall accuracy. Inception v3 and VGG19 also exhibited high accuracy. In the hold-out validation method, Resnet50, Inception v3, and Densenet121 showed high performance.

The conventional methods used for a definitive diagnosis of AD, such as genetic tests, cerebrospinal fluid tests, PET, and MR imaging, are invasive and costly, making them less accessible [23,24]. Moreover, these methods typically require expert interpretation [25,26]. In contrast, the approach presented in this study, utilizing a deep learning model based on voice analysis, offers a non-invasive, cost-effective, and time-efficient alternative. If it is more systematized, it may be automated and utilized without the help of experts.

Overall, the Densenet121, Inception v3, resnet50, and VGG19 models performed excellently and exhibited similar or superior performance to the MMSE-DS [59]. When the voice signals of the responses were analyzed to classify the AD patients, we achieved a sensitivity of 0.9550, a specificity of 0.8333, an accuracy of 0.9000, and an area under the curve (AUC) of 0.9243 in the five-fold cross-validation method. The accuracy of this study, wherein the spectrogram of the voice data was used to train a convolutional neural network (CNN) to classify AD patients, was higher than that reported by Duc and Ryu (85.27%), who investigated the correlation between the 3D-functional MRI results and MMSE scores [78], and Tae Hui Kim (0.895), who evaluated the diagnostic accuracy of MMSE-DS [59]. Additionally, this study showed greater accuracy than the results reported by Liu and Cheng (91.2%), who used FDG-PET images to classify AD patients using a CNN and a recurrent neural network [79].

Considering AD patients were classified solely based on the analysis of voice signals, questions from the MMSE-DS requiring execution, visuospatial construction, judgment, and abstract thinking were not included in the experimental data based on recommendations from psychiatrists. This study aimed to show that AD classification is possible without these elements, and hence, the results are significant. In addition, it is considered meaningful to confirm which questionnaire among the MMSE-DS is suitable for voice-based deep learning classification. The academia community related to AD in South Korea judged that these results were reliable and useful for screening AD patients. In this study, an individual effect analysis of each question was not performed. It is expected that more diverse and accurate results can be acquired if the scope is expanded by performing an individual analysis of each question, a validation of a combination of various questions, and an analysis including additional elements.

We have tried to detect Alzheimer's disease based on the audio data without including any semantics. Therefore, we focused on non-verbal characteristics in the responses of patients with Alzheimer's disease, and for this purpose, we recruited subjects and designed an experiment. Actually, the experiment's results showed that Alzheimer's dementia patients had distinctive features in intonation and nuance, as well as inaccurate pronunciation, a slow pace, and the elongation of vowel sounds, regardless of whether the answer was correct or not [71–74].

As this was an experimental study in which real AD patients were recruited, there are limitations in terms of the subjects and data. Because the experiment was conducted with a limited number of 80 people, this study could not perform external validation. In future studies, subject recruitment will be conducted on a large scale, and subjects from external institutions for external validation will be considered. In addition, other voice data, not responses to specific answers, will be collected, compared, and analyzed. Also, the severity of dementia will be classified using a deep learning model.

With respect to data augmentation, there are concerns that traditional data augmentation methods (flip, crop, enlargement, reduction, rotation, inversion, etc.) may negatively impact patterns or textures reflecting non-verbal characteristics in MFCC images. As can be seen in Tables A1 and A2, the augmentation range of the brightness and horizontal shift also preserves the quality of usable data by validating the effective range. It is expected that more accurate and valid results can be obtained by increasing the number of participating subjects and the amount of data.

From a clinical point of view, our results have several strengths. First, MMSE-DS requires checking all 28 items, while our results require only 12 items, making it simple and time-saving. Second, the MMSE-DS should have different cut-off scores according to educational background, age, and gender. It is convenient and useful because clinicians do not have to consider various conditions for dementia screening.

But there are also limitations. Dementia occurs when there is a severe decrease in function in some of the various cognitive domains. Since this study used only temporal orientation, memory registration, delayed recall, and attention, the evaluation of other cognitive functions may be limited. However, as mentioned earlier, in the early stages of dementia, generally there is a decrease in peripheral awareness, a lack of concentration, distraction, memory impairment, and a decrease in language ability. In other words, if early dementia patients can be selected through voice and responses to limited items, it is expected to bring about a fundamental innovation in the dementia screening method.

This study is a simplified dementia screening test, so further validation of each question's reliability is needed, but it is considered to be a useful technique for screening high-risk groups. It will also be of great help in developing a platform that performs high-risk screening, precise diagnostic testing, and management.

If the deep learning model proposed in this study is used, AD screening can be performed more easily and quickly. Based on this, it is possible to build a telemedicine or screening automation system through smart devices [46,52]. By installing them in an easily accessible place, the barrier to entry can be lowered, and the quality of telemedicine can be improved by utilizing virtual reality [47–49]. In addition, since early diagnosis of AD can inhibit its progression, it is expected that the simple screening method introduced in this paper will also work as a digital therapeutic.

5. Conclusions

In this study, voice data were acquired using MMSE-DS to distinguish between AD patients and healthy adults with high accuracy. In the five-fold cross and hold-out validation, Densenet121, Inception v3, and Resnet50 deep learning models showed high performance with sensitivity, accuracy, PPV, F1-score, and AUC metrics. Their performance was higher than the classification accuracy of MMSE-DS, and they also recorded good results compared to other studies that did not use voice-deep learning. Utilizing the results of this study, the screening process for AD patients can be simplified, which can contribute to increasing the accessibility of AD testing and the early diagnosis rate. In addition, it can be developed into an automated system to reduce dependence on experts and can contribute to AD screening by being applied to remote or online medical treatment.

Author Contributions: K.A., M.C., S.W.K., K.E.L. and H.-J.K., conceptualization; K.A. and M.C., formal analysis; K.A., M.C. and S.Y., software; K.A., M.C. and S.Y. validation; K.A., M.C. and Y.S., writing; S.W.K., K.E.L. and Y.S., data duration; K.A. and M.C., writing—original draft preparation; Y.S. and H.-J.K., writing—review and editing; S.W.K., K.E.L., S.Y.J., J.L.K. and D.H.Y., supervision; K.A., M.C. and H.-J.K., project administration; S.Y.J. and J.L.K., resources; K.A. and M.C. and H.-J.K., funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Institute of Information and Communications Technology Planning and Evaluation (IITP) grant funded by the Korea Government (MSIT) (No. 2021-0-00312).

Institutional Review Board Statement: This study was approved by the institutional review board of Chungnam National University Hospital (human subject study, prospective study, observational study, controlled study) (IRB approval number: CNUH2019-02-068), and the study protocol adhered to the ethical guidelines of the 1975 Declaration of Helsinki.

Informed Consent Statement: Informed consent was obtained from all patients involved in the study.

Data Availability Statement: Data was evaluated in the process of the manuscript reviewing (https://github.com/mikeahn00/ahn_experiment), but I concerned the data open access due to privacy.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

AD, Alzheimer’s disease; AUC, area under curve; CNN, convolutional neural network; MFCC, Mel-frequency cepstral coefficient; MMSE-DS, mini-mental state examination for dementia screening; MRI, magnetic resonance imaging; NPV, negative predictive values; PET, positron emission tomography; PPV, positive predictive values; ROC, receiver operating characteristic; t-SNE, t-distributed stochastic neighbor embedding.

Appendix A

Table A1. The augmentation range of the brightness.

Model Name	Augmentation	Sensitivity	Specificity	Accuracy	PPV	NPV	F1-Score	AUC
Densnet121	Brightness $\pm 5\%$	1.0000	0.8571	0.9375	0.9000	1.0000	0.9473	1.0000
Densnet121	Brightness $\pm 10\%$	1.0000	0.8000	0.8750	0.7500	1.0000	0.8571	0.9666
Densnet121	Brightness $\pm 15\%$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9682
Densnet121	Brightness $\pm 20\%$	1.0000	0.8888	0.9375	0.8750	1.0000	0.9333	0.9523

Table A2. The augmentation range of the horizontal shift.

Model Name	Augmentation	Sensitivity	Specificity	Accuracy	PPV	NPV	F1-Score	AUC
Densnet121	Right shift 5%	1.0000	0.8750	0.9375	0.8888	1.0000	0.9411	0.9375
Densnet121	Right shift 10%	1.0000	0.8571	0.9375	0.9000	1.0000	0.9473	0.9365
Densnet121	Right shift 15%	1.0000	0.9000	0.9375	0.8571	1.0000	0.9230	0.9833
Densnet121	Right shift 20%	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Densnet121	Right shift 25%	0.8571	1.0000	0.9375	1.0000	0.9000	0.9230	1.0000

The confusion matrix for each algorithm is in Figures A1 and A2.

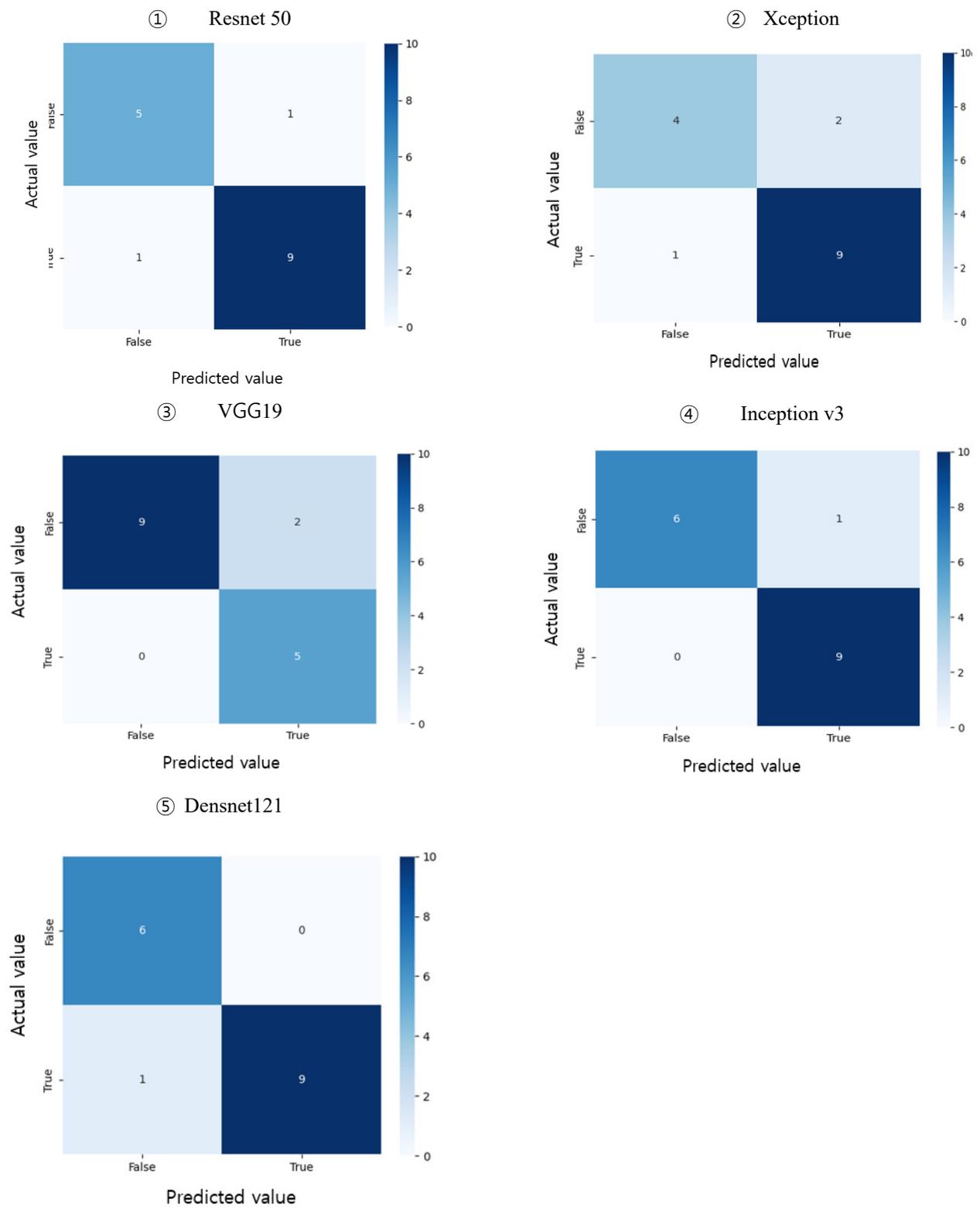


Figure A1. The confusion matrix for each algorithm of the Hold-out method.

① Resnet50

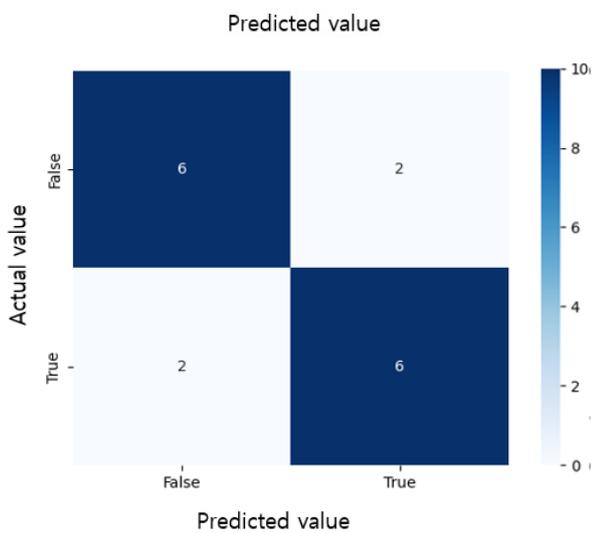
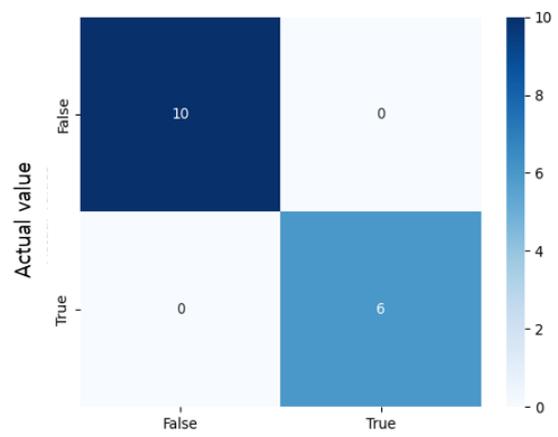
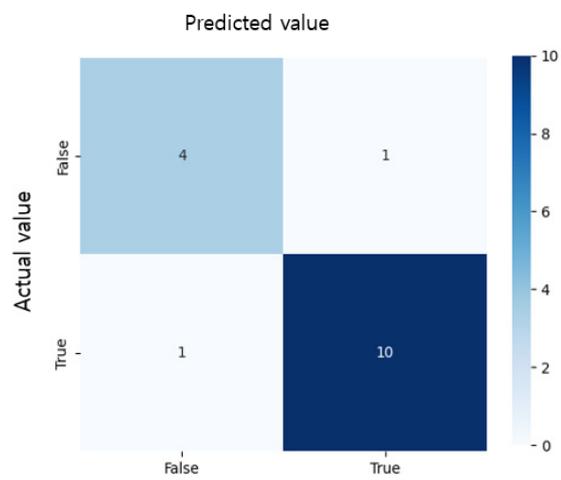
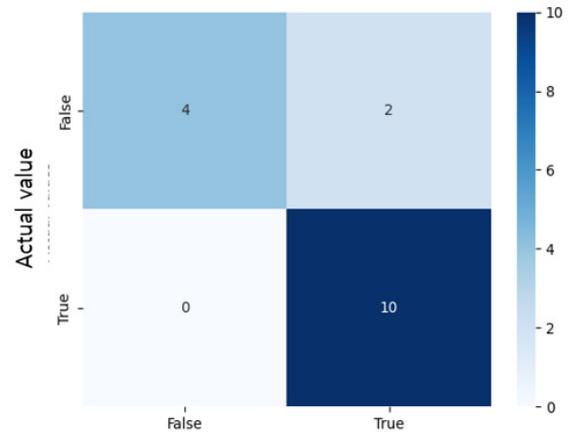
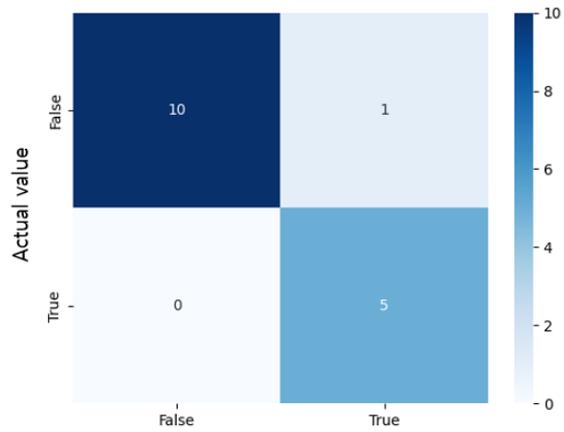


Figure A2. Cont.

② Xception

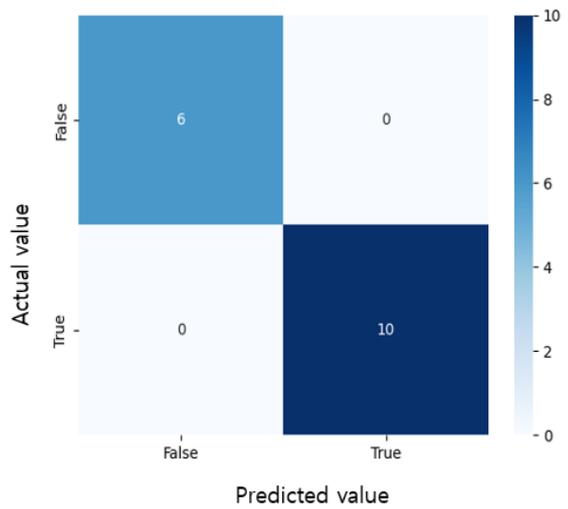
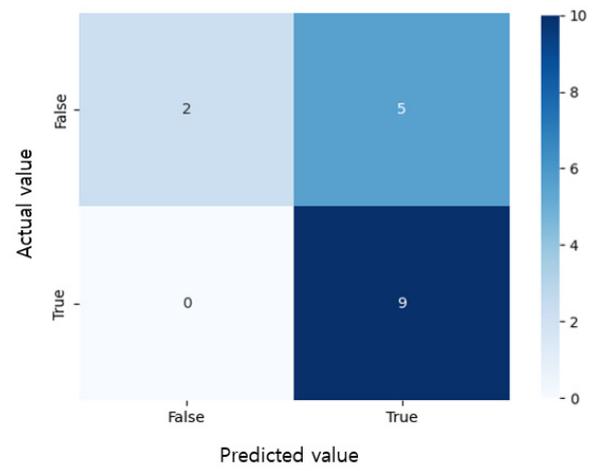
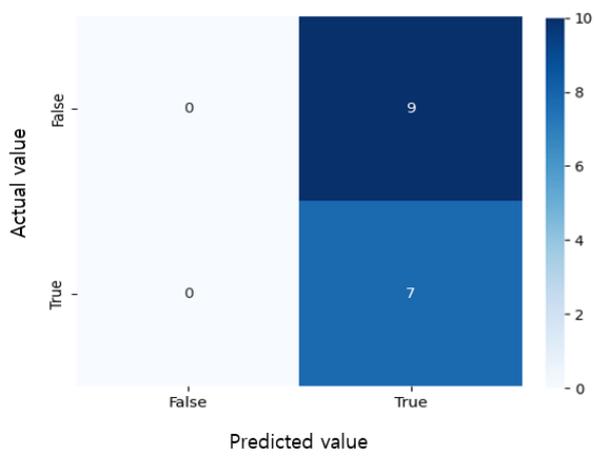
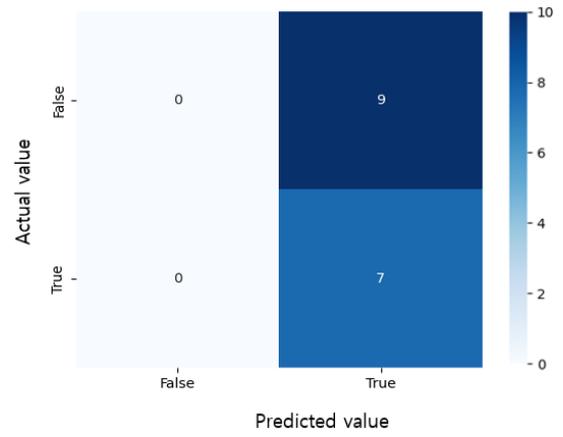
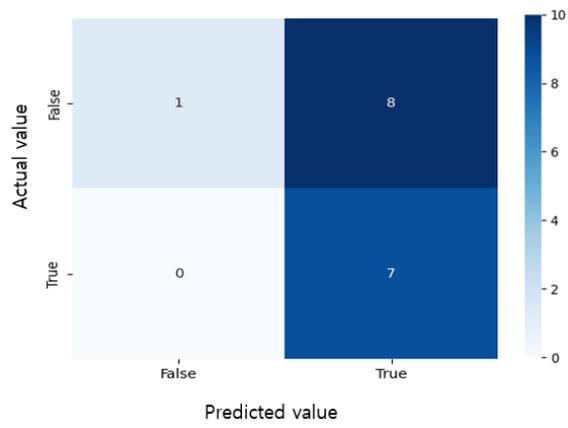


Figure A2. Cont.

③ VGG19

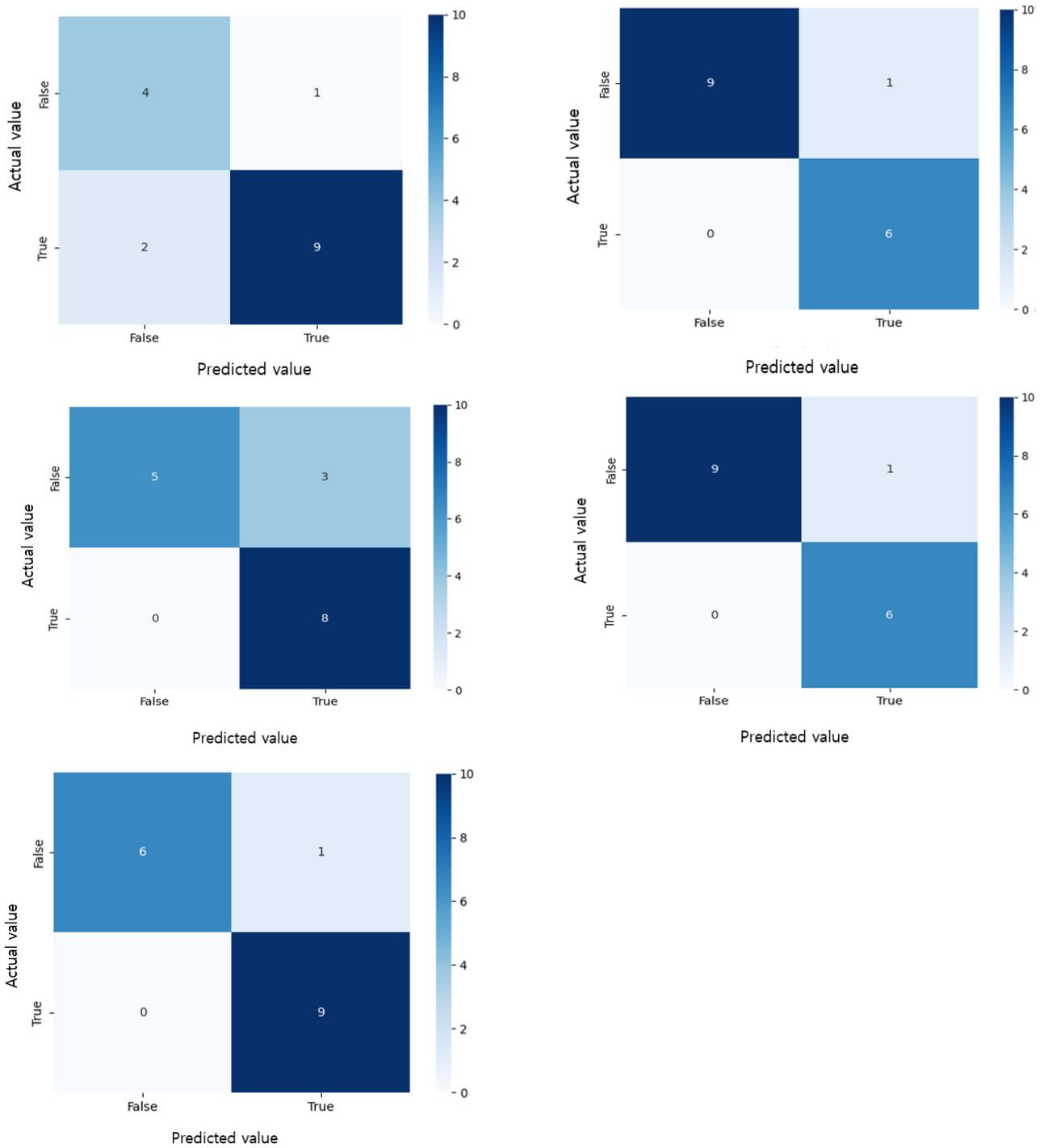


Figure A2. Cont.

④ Inception v3

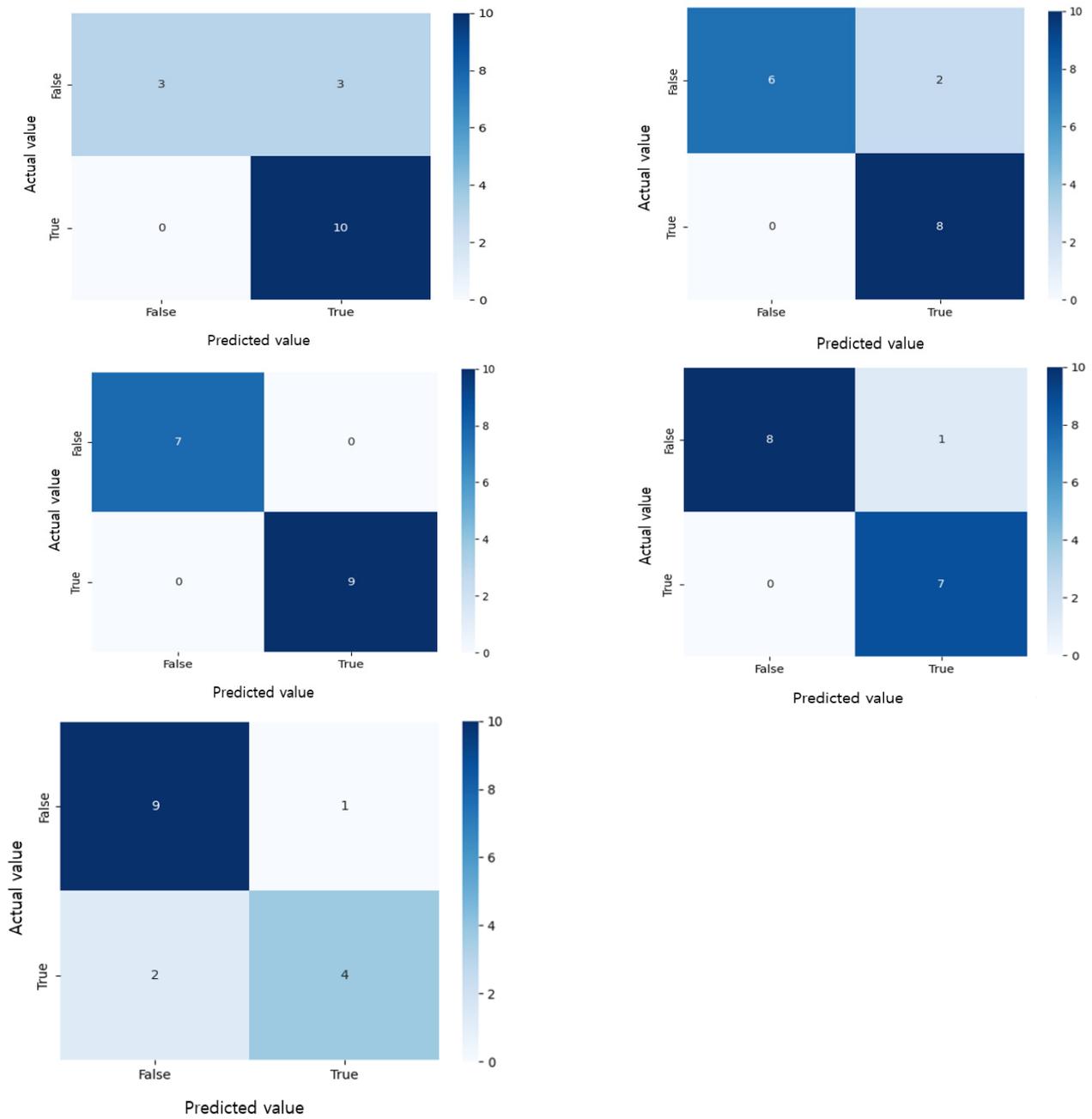


Figure A2. Cont.

⑤ Densnet121

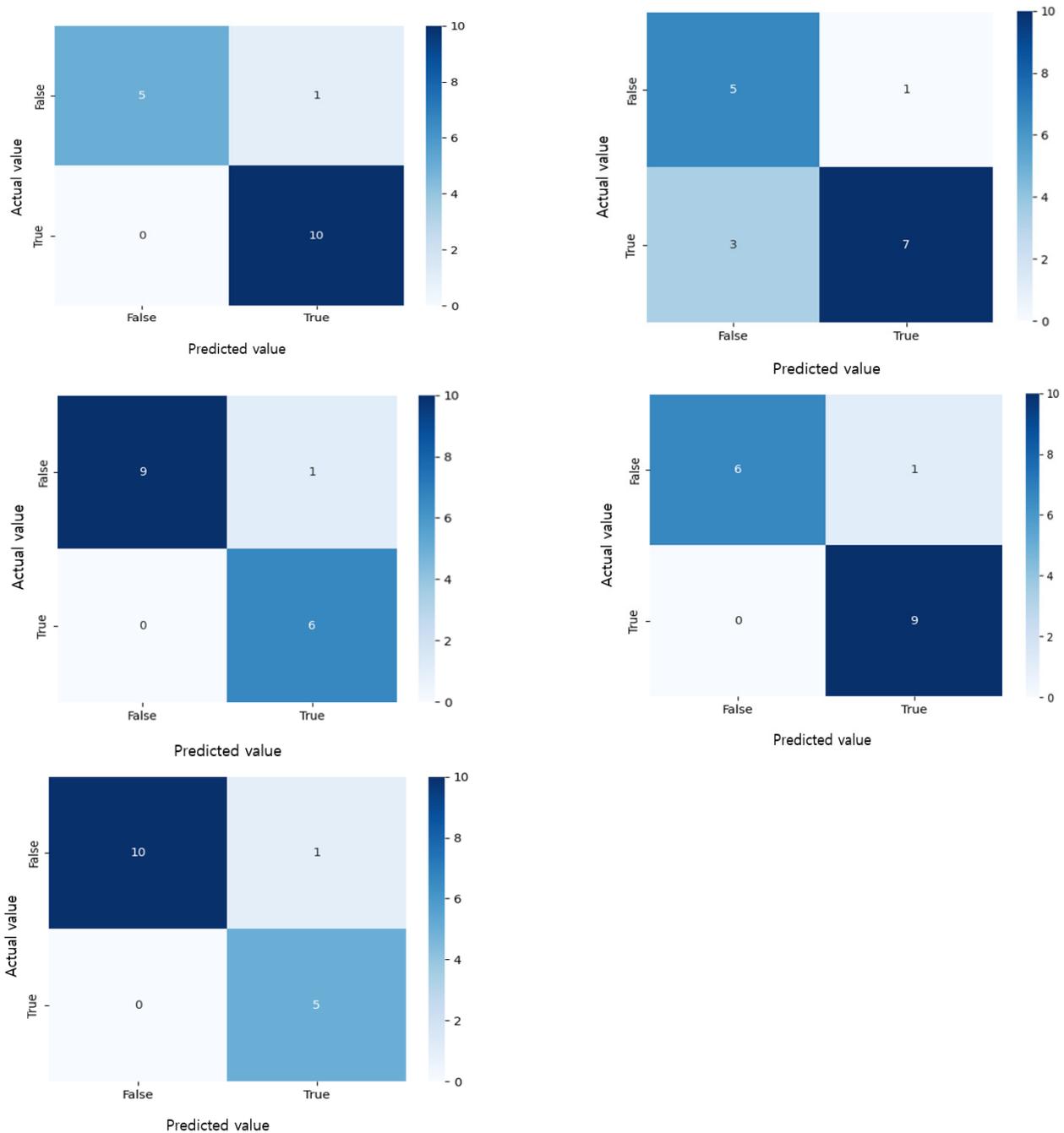


Figure A2. The confusion matrix for each algorithm of the Five-fold method.

References

1. Palop, J.J.; Chin, J.; Mucke, L. A network dysfunction perspective on neurodegenerative diseases. *Nature* **2006**, *443*, 768–773. [CrossRef]
2. McGeer, P.L.; McGeer, E.G. The inflammatory response system of brain: Implications for therapy of Alzheimer and other neurodegenerative diseases. *Brain Res. Rev.* **1995**, *21*, 195–218. [CrossRef]
3. Cho, E.; Park, M. Palmitoylation in Alzheimer’s disease and other neurodegenerative diseases. *Pharmacol. Res.* **2016**, *111*, 133–151. [CrossRef] [PubMed]
4. Huang, L.-K.; Chao, S.-P.; Hu, C.-J. Clinical trials of new drugs for Alzheimer disease. *J. Biomed. Sci.* **2020**, *27*, 18. [CrossRef]
5. Zanetti, O.; Solerte, S.B.; Cantoni, F. Life expectancy in Alzheimer’s disease (AD). *Arch. Gerontol. Geriatr.* **2009**, *49*, 237–243. [CrossRef]

6. Patwardhan, A.G.; Belemkar, S. An update on Alzheimer's disease: Immunotherapeutic agents, stem cell therapy and gene editing. *Life Sci.* **2021**, *282*, 119790. [[CrossRef](#)] [[PubMed](#)]
7. Janoutová, J.; Kovalová, M.; Machaczka, O.; Ambroz, P.; Zatloukalová, A.; Němček, K.; Janout, V. Risk Factors for Alzheimer's Disease: An Epidemiological Study. *Curr. Alzheimer Res.* **2021**, *18*, 372–379. [[CrossRef](#)]
8. Kuang, J.; Zhang, P.; Cai, T.; Zou, Z.; Li, L.; Wang, N.; Wu, L. Prediction of transition from mild cognitive impairment to Alzheimer's disease based on a logistic regression–artificial neural network–decision tree model. *Geriatr. Gerontol. Int.* **2020**, *21*, 43–47. [[CrossRef](#)]
9. Chen, S.; Song, Y.; Xu, W.; Hu, G.; Ge, H.; Xue, C.; Gao, J.; Qi, W.; Lin, X.; Chen, J.; et al. Impaired Memory Awareness and Loss Integration in Self-Referential Network Across the Progression of Alzheimer's Disease Spectrum. *J. Alzheimer's Dis.* **2021**, *83*, 111–126. [[CrossRef](#)]
10. Chyniak, O.S.; Dubenko, O.Y.; Potapov, O.O. The relationship between decreased cognitive functions and the level of proinflammatory cytokines in patients with Alzheimer's disease, vascular dementia, and mild cognitive disorder. *Eastern Ukr. Med. J.* **2021**, *9*, 247–255.
11. Bavarsad, K.; Hosseini, M.; Hadjzadeh, M.; Sahebkar, A. The effects of thyroid hormones on memory impairment and Alzheimer's disease. *J. Cell. Physiol.* **2019**, *234*, 14633–14640. [[CrossRef](#)]
12. Perry, R.J.; Watson, P.; Hodges, J.R. The nature and staging of attention dysfunction in early (minimal and mild) Alzheimer's disease: Relationship to episodic and semantic memory impairment. *Neuropsychologia* **1999**, *38*, 252–271. [[CrossRef](#)] [[PubMed](#)]
13. Price, B.H.; Gurvit, H.; Weintraub, S.; Geula, C.; Leimkuhler, E.; Mesulam, M. Neuropsychological Patterns and Language Deficits in 20 Consecutive Cases of Autopsy-Confirmed Alzheimer's Disease. *Arch. Neurol.* **1993**, *50*, 931–937. [[CrossRef](#)] [[PubMed](#)]
14. Sarazin, M.; Stern, Y.; Berr, C.; Riba, A.; Albert, M.; Brandt, J.; Dubois, B. Neuropsychological predictors of dependency in patients with Alzheimer disease. *Neurology* **2005**, *64*, 1027–1031. [[CrossRef](#)]
15. Caramelli, P.; Mansur, L.L.; Nitri, R. Language and Communication Disorders in Dementia of the Alzheimer Type. In *Handbook of Neurolinguistics*; Academic Press: New York, NY, USA, 1998; pp. 463–473. [[CrossRef](#)]
16. Mackenzie, T.B.; Robiner, W.N.; Knopman, D.S. Differences between patient and family assessments of depression in Alzheimer's disease. *Am. J. Psychiatry* **1989**, *146*, 1174–1178. [[CrossRef](#)]
17. Sheth, K. Alzheimer's' The Family Disease': Examining the Effects of Resilience on Preparedness and Compassion in Asian Family Caregivers. Ph.D. Thesis, Alliant International University, San Diego, CA, USA, 2020.
18. Zvěřová, M. Clinical aspects of Alzheimer's disease. *Clin. Biochem.* **2019**, *72*, 3–6. [[CrossRef](#)]
19. Banerjee, S. The macroeconomics of dementia—Will the world economy get Alzheimer's disease? *Arch. Med. Res.* **2012**, *43*, 705–709. [[CrossRef](#)]
20. Thorpe, K.E.; Ogden, L.L.; Galactionova, K. Chronic conditions account for rise in medicare spending from 1987 to 2006. *Health Aff.* **2010**, *29*, 718–724. [[CrossRef](#)] [[PubMed](#)]
21. Nestor, P.J.; Scheltens, P.; Hodges, J.R. Advances in the early detection of Alzheimer's disease. *Nat. Med.* **2004**, *10*, S34–S41. [[CrossRef](#)]
22. Welsh, K.; Butters, N.; Hughes, J.; Mohs, R.; Heyman, A. Detection of Abnormal Memory Decline in Mild Cases of Alzheimer's Disease Using CERAD Neuropsychological Measures. *Arch. Neurol.* **1991**, *48*, 278–281. [[CrossRef](#)]
23. Bharati, S.; Podder, P.; Thanh, D.N.H.; Prasath, V.B.S. Dementia classification using MR imaging and clinical data with voting based machine learning models. *Multimed. Tools Appl.* **2022**, *81*, 25971–25992. [[CrossRef](#)]
24. McGeer, P.L.; Kamo, H.; Harrop, R.; McGeer, E.G.; Martin WR, W.; Pate, B.D.; Li DK, B. Comparison of PET, MRI, and CT with pathology in a proven case of Alzheimer's disease. *Neurology* **1986**, *36*, 1569. [[CrossRef](#)]
25. Weiner, M.F. Alzheimer's Disease: Diagnosis and Treatment. *Harv. Rev. Psychiatry* **1997**, *4*, 306–316. [[CrossRef](#)] [[PubMed](#)]
26. Jha, A.; Mukhopadhyaya, K. *Alzheimer's Disease: Diagnosis and Treatment Guide*; Springer Nature: Berlin/Heidelberg, Germany, 2020.
27. Kato, S.; Homma, A.; Sakuma, T. Easy Screening for Mild Alzheimer's Disease and Mild Cognitive Impairment from Elderly Speech. *Curr. Alzheimer Res.* **2018**, *15*, 104–110. [[CrossRef](#)] [[PubMed](#)]
28. Zhou, X.; Ashford, J.W. Advances in screening instruments for Alzheimer's disease. *Aging Med.* **2019**, *2*, 88–93. [[CrossRef](#)]
29. De Roeck, E.E.; De Deyn, P.P.; Dierckx, E.; Engelborghs, S. Brief cognitive screening instruments for early detection of Alzheimer's disease: A systematic review. *Alzheimer's Res. Ther.* **2019**, *11*, 21. [[CrossRef](#)]
30. Monsch, A.U.; Bondi, M.W.; Butters, N.; Salmon, D.P.; Katzman, R.; Thal, L.J. Comparisons of Verbal Fluency Tasks in the Detection of Dementia of the Alzheimer Type. *Arch. Neurol.* **1992**, *49*, 1253–1258. [[CrossRef](#)]
31. Cuetos, F.; Arango-Lasprilla, J.C.; Uribe, C.; Valencia, C.; Lopera, F. Linguistic changes in verbal expression: A preclinical marker of Alzheimer's disease. *J. Int. Neuropsychol. Soc.* **2007**, *13*, 433–439. [[CrossRef](#)]
32. Taler, V.; Phillips, N.A. Language performance in Alzheimer's disease and mild cognitive impairment: A comparative review. *J. Clin. Exp. Neuropsychol.* **2008**, *30*, 501–556. [[CrossRef](#)]
33. Reilly, J.; Rodriguez, A.D.; Lamy, M.; Neils-Strunjas, J. Cognition, language, and clinical pathological features of non-Alzheimer's dementias: An overview. *J. Commun. Disord.* **2010**, *43*, 438–452. [[CrossRef](#)]

34. Lopez-De-Ipina, K.; Martinez-De-Lizarduy, U.; Calvo, P.M.; Mekyska, J.; Beitia, B.; Barroso, N.; Estanga, A.; Tainta, M.; Ecay-Torres, M. Advances on Automatic Speech Analysis for Early Detection of Alzheimer Disease: A Non-linear Multi-task Approach. *Curr. Alzheimer Res.* **2018**, *15*, 139–148. [[CrossRef](#)]
35. Meilan, J.J.; Martinez-Sanchez, F.; Carro, J.; Carcavilla, N.; Ivanova, O. Voice Markers of Lexical Access in Mild Cognitive Impairment and Alzheimer's Disease. *Curr. Alzheimer Res.* **2018**, *15*, 111–119. [[CrossRef](#)]
36. Martínez-Nicolás, I.; Llorente, T.E.; Martínez-Sánchez, F.; Meilán, J.J.G. Ten Years of Research on Automatic Voice and Speech Analysis of People With Alzheimer's Disease and Mild Cognitive Impairment: A Systematic Review Article. *Front. Psychol.* **2021**, *12*, 620251. [[CrossRef](#)]
37. Ortega, L.V.; Aprahamian, I.; Martinelli, J.E.; Cecchini, M.A.; Cação, J.d.C.; Yassuda, M.S. Diagnostic Accuracy of Usual Cognitive Screening Tests Versus Appropriate Tests for Lower Education to Identify Alzheimer Disease. *J. Geriatr. Psychiatry Neurol.* **2020**, *34*, 222–231. [[CrossRef](#)]
38. Amlerova, J.; Laczó, J.; Nedelska, Z.; Laczó, M.; Vyhnálek, M.; Zhang, B.; Sheardova, K.; Angelucci, F.; Andel, R.; Hort, J. Emotional prosody recognition is impaired in Alzheimer's disease. *Alzheimer's Res. Ther.* **2022**, *14*, 50. [[CrossRef](#)]
39. Oh, C.; Morris, R.J.; Wang, X. A Systematic Review of Expressive and Receptive Prosody in People With Dementia. *J. Speech Lang. Hear. Res.* **2021**, *64*, 3803–3825. [[CrossRef](#)]
40. Tosto, G.; Gasparini, M.; Lenzi, G.L.; Bruno, G. Prosodic impairment in Alzheimer's disease: Assessment and clinical relevance. *J. Neuropsychiatry Clin. Neurosci.* **2011**, *23*, E21–E23. [[CrossRef](#)]
41. Kundaram, S.S.; Pathak, K.C. Deep learning-based alzheimer disease detection. In Proceedings of the Fourth International Conference on Microelectronics, Computing and Communication Systems, Ranchi, India, 11–12 May 2019; Springer: Singapore, 2021.
42. Pathak, K.C.; Kundaram, S.S. Accuracy-Based Performance Analysis of Alzheimer's Disease Classification Using Deep Convolution Neural Network. In *Soft Computing: Theories and Applications: Proceedings of SoCTA 2019*; Springer: Singapore, 2020; pp. 731–744. [[CrossRef](#)]
43. Takeda, C.; Guyonnet, S.; Ousset, P.; Soto, M.; Vellas, B. Toulouse Alzheimer's Clinical Research Center recovery after the COVID-19 crisis: Telemedicine an innovative solution for clinical research during the coronavirus pandemic. *J. Prev. Alzheimer's Dis.* **2020**, *7*, 301–304. [[CrossRef](#)]
44. Capozzo, R.; Zoccolella, S.; Frisullo, M.E.; Barone, R.; Dell'abate, M.T.; Barulli, M.R.; Musio, M.; Accogli, M.; Logroscino, G. Telemedicine for Delivery of Care in Frontotemporal Lobar Degeneration during COVID-19 Pandemic: Results from Southern Italy. *J. Alzheimer's Dis.* **2020**, *76*, 481–489. [[CrossRef](#)]
45. Nester, C.O.; Mogle, J.; Katz, M.J.; Wang, C.; Lipton, R.B.; Derby, C.A.; Rabin, L. Aging research in the time of COVID-19: A telephone screen for subjective cognitive concerns in community-dwelling ethnically diverse older adults. *Alzheimer's Dement.* **2021**, *17*, e056403. [[CrossRef](#)]
46. Sotaniemi, M.; Pulliainen, V.; Hokkanen, L.; Pirttilä, T.; Hallikainen, I.; Soinen, H.; Hänninen, T. CERAD-neuropsychological battery in screening mild Alzheimer's disease. *Acta Neurol. Scand.* **2011**, *125*, 16–23. [[CrossRef](#)]
47. Gosse, P.J.; Kassardjian, C.D.; Masellis, M.; Mitchell, S.B. Virtual care for patients with Alzheimer disease and related dementias during the COVID-19 era and beyond. *Can. Med. Assoc. J.* **2021**, *193*, E371–E377. [[CrossRef](#)]
48. Matamala-Gomez, M.; Bottiroli, S.; Realdon, O.; Riva, G.; Galvagni, L.; Platz, T.; Sandrini, G.; De Icco, R.; Tassorelli, C. Telemedicine and Virtual Reality at Time of COVID-19 Pandemic: An Overview for Future Perspectives in Neurorehabilitation. *Front. Neurol.* **2021**, *12*, 646902. [[CrossRef](#)] [[PubMed](#)]
49. Mantovani, E.; Zucchella, C.; Bottiroli, S.; Federico, A.; Giugno, R.; Sandrini, G.; Chiamulera, C.; Tamburin, S. Telemedicine and Virtual Reality for Cognitive Rehabilitation: A Roadmap for the COVID-19 Pandemic. *Front. Neurol.* **2020**, *11*, 926. [[CrossRef](#)] [[PubMed](#)]
50. Prins, S.; Zhuparris, A.; Hart, E.P.; Doll, R.J.; Groeneveld, G.J. A cross-sectional study in healthy elderly subjects aimed at development of an algorithm to increase identification of Alzheimer pathology for the purpose of clinical trial participation. *Alzheimer's Res. Ther.* **2021**, *13*, 132. [[CrossRef](#)] [[PubMed](#)]
51. Sato, K.; Mano, T.; Ihara, R.; Suzuki, K.; Niimi, Y.; Toda, T.; Iwatsubo, T.; Iwata, A.; Alzheimer's Disease Neuroimaging Initiative; Japanese Alzheimer's Disease Neuroimaging Initiative; et al. Cohort-specific optimization of models predicting preclinical Alzheimer's disease, to enhance screening performance in the middle of preclinical Alzheimer's disease clinical studies. *J. Prev. Alzheimer's Dis.* **2021**, *8*, 503–512. [[CrossRef](#)]
52. Kwon, S.J.; Kim, H.S.; Han, J.H.; Bin Bae, J.; Kim, K.W. Reliability and Validity of Alzheimer's Disease Screening With a Semi-automated Smartphone Application Using Verbal Fluency. *Front. Neurol.* **2021**, *12*, 684902. [[CrossRef](#)]
53. Bae, M.; Chang, K.J. Cognitive function of the elderly without dementia in Korea: Cross-sectional study. *Alzheimer's Dement.* **2021**, *17*, e051714. [[CrossRef](#)]
54. Jung, Y.-S.; Park, T.; Kim, E.-K.; Jeong, S.-H.; Lee, Y.-E.; Cho, M.-J.; Song, K.-B.; Choi, Y.-H. Influence of Chewing Ability on Elderly Adults' Cognitive Functioning: The Mediating Effects of the Ability to Perform Daily Life Activities and Nutritional Status. *Int. J. Environ. Res. Public Health* **2022**, *19*, 1236. [[CrossRef](#)]
55. Suh, H.-W.; Seol, J.-H.; Bae, E.-J.; Kwak, H.-Y.; Hong, S.; Park, Y.-S.; Lim, J.H.; Chung, S.-Y. Effectiveness and Safety of the Korean Medicine Senior Health Promotion Program Using Herbal Medicine and Acupuncture for Mild Cognitive Impairment: A Retrospective Study of 500 Patients in Seoul, Korea. *Evid.-Based Complement. Altern. Med.* **2021**, *2021*, 8820705. [[CrossRef](#)]

56. Park, Y.-S.; Jee, Y.-J. The Effects on Cognitive, Emotional, and Physical Functions of the Elderly at Local Senior Center from Dementia Prevention Program for the Dementia Safety Village. *J. Med. Imaging Health Inform.* **2021**, *11*, 508–512. [[CrossRef](#)]
57. Kim, M.; Kim, H.; Lim, J.S. Classification of Diagnosis of Alzheimer's Disease Based on Convolutional Layers of VGG16 Model using Speech Data. In Proceedings of the 2020 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, Republic of Korea, 21–23 October 2020; IEEE: Piscataway, NJ, USA, 2020. [[CrossRef](#)]
58. Ha, J.-H.; Kwak, M.; Han, J.W.; Lee, H.J.; Ingersoll-Dayton, B.; Spencer, B.; Kim, K.W. The effectiveness of a couple-based intervention for people living with mild Alzheimer's disease and their spousal caregivers in Korea. *Dementia* **2018**, *20*, 831–847. [[CrossRef](#)] [[PubMed](#)]
59. Kim, T.H.; Jhoo, J.H.; Park, J.H.; Kim, J.L.; Ryu, S.H.; Moon, S.W.; Choo, I.H.; Lee, D.W.; Yoon, J.C.; Do, Y.J.; et al. Korean Version of Mini Mental Status Examination for Dementia Screening and Its' Short Form. *Psychiatry Investig.* **2010**, *7*, 102–108. [[CrossRef](#)]
60. HAMMINJO; Lee, J.; Kim, S.; Yoo, D. The effects of a multimodal interventional program on cognitive function, instrumental activities of daily living in patients with mild Alzheimer's disease. *Korean J. Occup. Ther.* **2018**, *26*, 91–102.
61. Meghanani, A.; Anoop, C.S.; Ramakrishnan, A.G. An Exploration of Log-Mel Spectrogram and MFCC Features for Alzheimer's Dementia Recognition from Spontaneous Speech. In Proceedings of the 2021 IEEE Spoken Language Technology Workshop (SLT), Shenzhen, China, 19–22 January 2021; IEEE: Piscataway, NJ, USA, 2021. [[CrossRef](#)]
62. Mirzaei, S.; El Yacoubi, M.; Garcia-Salicetti, S.; Boudy, J.; Kahindo, C.; Cristancho-Lacroix, V.; Kerhervé, H.; Rigaud, A.-S. Two-Stage Feature Selection of Voice Parameters for Early Alzheimer's Disease Prediction. *IRBM* **2018**, *39*, 430–435. [[CrossRef](#)]
63. Dash, T.K.; Mishra, S.; Panda, G.; Satapathy, S.C. Detection of COVID-19 from speech signal using bio-inspired based cepstral features. *Pattern Recognit.* **2021**, *117*, 107999. [[CrossRef](#)]
64. Huang, L.; Pun, C.M. Audio replay spoof attack detection using segment-based hybrid feature and densenet-LSTM network. In Proceedings of the ICASSP 2019-2019 IEEE international conference on acoustics, speech and signal processing (ICASSP), Brighton, UK, 12–17 May 2019; IEEE: Piscataway, NJ, USA, 2019.
65. Doshi, S.; Patidar, T.; Gautam, S.; Kumar, R. *Acoustic Scene Analysis and Classification Using Densenet Convolutional Neural Network*; No. 8056; EasyChair: Stockport, UK, 2022.
66. Cui, S.; Huang, B.; Huang, J.; Kang, X. Synthetic Speech Detection Based on Local Autoregression and Variance Statistics. *IEEE Signal Process. Lett.* **2022**, *29*, 1462–1466. [[CrossRef](#)]
67. Wang, P.; Li, Y.; Vasconcelos, N. Rethinking and improving the robustness of image style transfer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021.
68. Pandhi, T.; Kapoor, T.; Gupta, B. An Improved Technique for Preliminary Diagnosis of COVID-19 via Cough Audio Analysis. In Proceedings of the International Conference on Recent Trends in Image Processing and Pattern Recognition, Msida, Malta, 8–10 December 2021; Springer: Cham, Switzerland, 2022.
69. Malov, D.; Shumskaya, O. Fatigue recognition based on audiovisual content. In Proceedings of the 2019 1st International Conference on Control Systems, Mathematical Modelling, Automation and Energy Efficiency (SUMMA), Lipetsk, Russia, 20–22 November 2019; IEEE: Piscataway, NJ, USA, 2019.
70. Ouyang, D.; Theurer, J.; Stein, N.R.; Hughes, J.W.; Elias, P.; He, B.; Yuan, N.; Duffy, G.; Sandhu, R.K.; Ebinger, J.; et al. Electrocardiographic Deep Learning for Predicting Post-Procedural Mortality. *arXiv* **2022**, arXiv:2205.03242.
71. Carpenter, C.R.; Bassett, E.R.; Fischer, G.M.; Shirshakan, J.; Galvin, J.E.; Morris, J.C. Four sensitive screening tools to detect cognitive dysfunction in geriatric emergency department patients: Brief Alzheimer's Screen, Short Blessed Test, Ottawa 3DY, and the Caregiver-Completed AD8. *Acad. Emerg. Med.* **2011**, *18*, 374–384. [[CrossRef](#)]
72. So, J.-H.; Madusanka, N.; Choi, H.-K.; Choi, B.-K.; Park, H.-G. Deep Learning for Alzheimer's Disease Classification using Texture Features. *Curr. Med. Imaging Former. Curr. Med. Imaging Rev.* **2019**, *15*, 689–698. [[CrossRef](#)]
73. Liu, Z.; Proctor, L.; Collier, P.; Casenhiser, D.; Paek, E.J.; Yoon, S.O.; Zhao, X. Machine learning of transcripts and audio recordings of spontaneous speech for diagnosis of Alzheimer's disease. *Alzheimer's Dement.* **2021**, *17*, e057556. [[CrossRef](#)]
74. Mittal, A.; Dua, M.; Dua, S. Classical and Deep Learning Data Processing Techniques for Speech and Speaker Recognitions. In *Deep Learning Approaches for Spoken and Natural Language Processing*; Springer: Cham, Switzerland, 2021; pp. 111–126. [[CrossRef](#)]
75. Mahmood, A.; Utku, K.Ö.S.E. Speech recognition based on convolutional neural networks and MFCC algorithm. *Adv. Artif. Intell. Res.* **2021**, *1*, 6–12.
76. Ayvaz, U.; Gürüler, H.; Khan, F.; Ahmed, N.; Whangbo, T.; Bobomirzaevich, A.A. Automatic Speaker Recognition Using Mel-Frequency Cepstral Coefficients Through Machine Learning. *Comput. Mater. Contin.* **2022**, *71*, 5511–5521. [[CrossRef](#)]
77. Mistry, D.S.; Kulkarni, A.V. Overview: Speech recognition technology, mel-frequency cepstral coefficients (mfcc), artificial neural network (ann). *Int. J. Eng. Res. Technol.* **2013**, *2*, 10.
78. Duc, N.T.; Ryu, S.; Qureshi, M.N.I.; Choi, M.; Lee, K.H.; Lee, B. 3D-Deep Learning Based Automatic Diagnosis of Alzheimer's Disease with Joint MMSE Prediction Using Resting-State fMRI. *Neuroinformatics* **2019**, *18*, 71–86. [[CrossRef](#)]
79. Liu, M.; Cheng, D.; Yan, W.; Alzheimer's Disease Neuroimaging Initiative. Classification of Alzheimer's Disease by Combination of Convolutional and Recurrent Neural Networks Using FDG-PET Images. *Front. Aging Neurosci.* **2018**, *12*, 35. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.