Convolutional Neural Networks to Classify Alzheimer's Disease Severity Based on SPECT Images: A Comparative Study

Wei-Chih Lien, Chung-Hsing Yeh, Chun-Yang Chang, Chien-Hsiang Chang, Chien-Hsu Chen, Yang-Cheng Lin

Submitted to: JMIR mHealth and uHealth
on: August 16, 2022

Table of Contents

Convolutional Neural Networks to Classify Alzheimer's Disease Severity Based on SPECT Images: A Comparative Study

Wei-Chih Lien[1]; Chung-Hsing Yeh[2]; Chun-Yang Chang[3]; Chien-Hsiang Chang[3]; Chien-Hsu Chen[3]; Yang-Cheng Lin[3]

[1]Department of Physical Medicine and Rehabilitation, National Cheng Kung University Hospital, College of Medicine National Cheng Kung University Tainan TW
[2]Faculty of Information Technology Monash University Melbourne AU
[3]Department of Industrial Design National Cheng Kung University Tainan TW

Corresponding Author:
Wei-Chih Lien
Department of Physical Medicine and Rehabilitation, National Cheng Kung University Hospital, College of Medicine
National Cheng Kung University
No. 138, Sheng Li Rd.
Tainan
TW

Abstract

Background: Image-based recognition has become a long-term topic in the field of artificial intelligence, and neuroimaging has gradually become a beneficial way to understand the course of Alzheimer's disease (AD).

Objective: The goal of this study is to compare the detection performance of convolutional neural networks (CNNs) on medical images to establish a classification model for epidemiological research. However, medical image analysis lacks large labeled training datasets, and thus many transfer learning-based methods have been proposed to solve few labels in the medical field.

Methods: Owing to the scarcity of image data from single-photon emission computed tomography (SPECT), this study uses transfer learning to compare the performance of diagnostic methods based on five different CNNs (two lightweight and three deeper-weight CNN models) to determine the most suitable model. Brain scan image data were collected from 36 male and 63 female subjects. This study used 4711 images as the input data for the model.

Results: The Cognitive Abilities Screening Instrument and Mini Mental State Exam scores of subjects with Clinical Dementia Rating (CDR) of 2 were significantly lower than those of subjects with CDR of 1 and 0.5. These results indicate that the ResNet model (the deeper-weight CNN model) exhibits the highest accuracy (70.79%) and can hence be used to improve the classification of mild cognitive impairment (MCI), mild AD, and moderate AD (CDRs of 0.5, 1, and 2, respectively).

Conclusions: This study successfully analyzes the classification performance of different CNN architectures for medical images and also proves the effectiveness of transfer learning in identifying the MCI, mild AD, and moderate AD scoring based on SPECT images.

(JMIR Preprints 16/08/2022:41921)
DOI: https://doi.org/10.2196/preprints.41921

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Convolutional Neural Networks to Classify Alzheimer's Disease Severity Based on SPECT Images: A Comparative Study

Wei-Chih Lien[1,2,*], Chung-Hsing Yeh[3], Chun-Yang Chang[4], Chien-Hsiang Chang[4], Chien-Hsu Chen[4], Yang-Cheng Lin[4,*]

[1]Department of Physical Medicine and Rehabilitation, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan
[2]Department of Physical Medicine and Rehabilitation, College of Medicine, National Cheng Kung University, Tainan, Taiwan
[3]Faculty of Information Technology, Monash University, Melbourne, Australia
[4]Department of Industrial Design, National Cheng Kung University, Tainan, Taiwan

[*]Correspondence: Wei-Chih Lien, Yang-Cheng Lin
No. 138, Sheng Li Road, Tainan, Taiwan; No. 1, University Road, Tainan, Taiwan
Tel +886-6-2353535 ext. 5240; +886-6-2757575 ext. 54328
Fax +886-6-2766106; +886-6-2746088
Email: lwclwhab@ms8.hinet.net, lyc0914@mail.ncku.edu.tw

Abstract

Purpose: Image-based recognition has become a long-term topic in the field of artificial intelligence, and neuroimaging has gradually become a beneficial way to understand the course of Alzheimer's disease (AD). The goal of this study is to compare the detection performance of convolutional neural networks (CNNs) on medical images to establish a classification model for epidemiological research. However, medical image analysis lacks large labeled training datasets, and thus many transfer learning-based methods have been proposed to solve few labels in the medical field.

Patients and methods: Owing to the scarcity of image data from single-photon emission computed tomography (SPECT), this study uses transfer learning to compare the performance of diagnostic methods based on five different CNNs (two lightweight and three deeper-weight CNN models) to determine the most suitable model. Brain scan image data were collected from 36 male and 63 female subjects. This study used 4711 images as the input data for the model.

Results: The Cognitive Abilities Screening Instrument and Mini Mental State Exam scores of subjects with Clinical Dementia Rating (CDR) of 2 were significantly lower than those of subjects with CDR of 1 and 0.5. These results indicate that the ResNet model (the deeper-weight CNN model) exhibits the highest accuracy (70.79%) and can hence be used to improve the classification of mild cognitive impairment (MCI), mild AD, and moderate AD (CDRs of 0.5, 1, and 2, respectively).

Conclusion: This study successfully analyzes the classification performance of different CNN architectures for medical images and also proves the effectiveness of transfer learning in identifying the MCI, mild AD, and moderate AD scoring based on

SPECT images.

Keywords: convolutional neural networks; Alzheimer's disease; single-photon emission computed tomography (SPECT); transfer learning; image recognition

Introduction

The Internet of things (IoT) has rapidly grown under the influence of the digital revolution in the past. With the realization of cloud services, everything can be connected to the Internet. In addition, almost all information is stored in a data-based form, providing users with a transparent and convenient way to collect and access large amounts of data in databases. With the advent of big data, the characteristics of volume, velocity, and variety assist in decision-making and process optimization. However, more than 80% of enterprise data are unstructured, growing at rates of 55% and 65% per year. Hence, it is difficult to analyze huge amounts of data with extremely high diversity in the traditional way.[1] Therefore, the primary direction of recent discussions and research is the extraction of meaningful information from these large amounts of data. In recent years, improvements in computer performance have led to the development of various data exploration methods that have been used in the fields of medicine, engineering, and management.[2] Neural networks (NNs) are among the fields that have received significant attention from the public. An NN is a series of algorithms that endeavors to recognize underlying relationships in a set of data through a process that mimics the behavior and function of the nervous system. It uses ample artificial neurons to simulate the ability of biological neurons to compute information.[3,4] An NN can perform parallel processing and fault tolerance and can understand the relationship between samples to solve many complex classification and prediction problems.[5–7] Subsequently, from a different perspective, convolutional neural networks (CNNs) have been developed.[8,9] Compared with NNs, CNNs are more suitable for image recognition, language processing, and other applications.[10–12] With advancements in computer technology, the time required for computing has been drastically reduced. Accordingly, CNNs have been successfully applied in real-life applications, including speech recognition, handwriting recognition, and disease prediction.[13–17]
Medical diagnosis is a typical classification problem. Physicians use the behavior and symptoms of a patient to determine results based on previous experiences. In medical diagnosis, the value of medical imaging is indispensable, and it is currently the main noninvasive diagnostic tool. Currently, 90% of medical data still require physicians to use images to determine a patient's pathology, surgical plan, and medication risk.[18] However, most clinical diagnoses are based on the physician's past experience, which is extremely subjective. As the amount of data increases, the diagnosis period becomes relatively extended, resulting in an increase in the probability of misdiagnosis.[19] Nowadays, the factors that cause these diseases are extremely diverse. If data mining technology can be used effectively, physicians can improve the diagnostic accuracy. Effectively utilizing resources from existing medical data to improve the overall system function is the most important topic in the current research.
With the advent of an aging society, the number of people afflicted with Alzheimer's disease (AD) is increasing. Globally, approximately 9.9 million people suffer from dementia each year, implying that a new patient is diagnosed with the disease every 3.2 s. The number of people affected is expected to exceed 100 million by 2050. However, an effective medical treatment for the contraction of the disease still does not exist. In recent years, neuroimaging has gradually become a useful method to understand the course of AD, and the importance of using deep learning in the field of medical imaging has made great progress in research and clinical medicine.[20] The diffuse brain atrophy is noted in the macroscopic findings of AD.[21] The classification accuracy of AD versus healthy control using deep learning of magnetic resonance

imaging (MRI) is 91.4%, and that of mild cognitive impairment (MCI) versus AD is 70.1%.[22] The economic evaluation of dynamic susceptibility MRI compared to non-enhanced computer-assisted tomography (CT) was 479,500 US dollars per quality-adjusted life years, while the corresponding comparison for single-photon emission computed tomography (SPECT) compared to CT was dominant (higher effectiveness and lower cost).[23] From the perspective of neuropathology, subjects with very mild AD typically show abnormal metabolic and regional cerebral blood flow (rCBF) patterns, even at the preclinical stage.[24] In addition to clinical manifestations, analysis via SPECT has been frequently used by physicians as a diagnostic tool.[25] In previous studies, rCBF SPECT for the diagnosis of AD revealed a sensitivity of 86%, specificity of 73%, and accuracy of 82%.[26] Brain perfusion SPECT has been proven to be a sensitive tool for assessing functional deficits in the early stages of AD.[27] MCI, referred to with a Clinical Dementia Rating (CDR) of 0.5, with a pooled sensitivity and specificity of 93% and 97%, respectively,[28] is an emerging entity for early detection and intervention. To the best of our knowledge, no study has used rCBF SPECT to classify MCI and mild and moderate AD.

Specifically, only a few studies have comprehensively analyzed various types of CNN models and discussed their applications in medical imaging.[29–31] The purpose of this study is to compare various methods of CNN model detection and medical diagnosis, from MCI to AD. Therefore, we applied lightweight CNN models, such as MobileNet V2[32] and NASNetMobile[33], and deeper-weight CNN models, such as VGG16[34], InceptionV3[35], and ResNet[36]. The five CNN models mentioned above were the most commonly used in transfer learning for disease diagnosis using medical imaging in previous studies,[37] and their detection performance was compared. The findings of this study are expected to assist physicians in the early stages of clinical diagnosis and reduce the occurrence of misdiagnoses.

Materials and methods

This study was approved by the National Cheng Kung University Human Research Ethics Committee (NCKU HREC-E-108-282-2) and conformed to the principles of the Declaration of Helsinki. All participants provided written informed consent prior to participating in the study. The experimental steps for the CNN model training in this study are divided into two stages. The experimental process is illustrated in Figure 1. This study aims to compare the different architectures of CNN models with SPECT images. In terms of the analysis strategy, adjustments were made to three influencing factors, i.e., optimizer, fully connected layers, and model parameters, to optimize the performance of the CNN model.

In the first stage of the experiment, the conditions for screening CNN architectures mainly include information security risks and embedded devices. First, to make the system convenient, we expect to apply the study results to embedded terminal devices and small mobile devices in hospitals in the future, which can be deeply integrated with the medical diagnosis process, and then promote the development of artificial intelligence (AI) in the medical field to improve the efficiency of diagnosis. However, owing to the limited storage space and power consumption of general devices, the feasibility of applying deep CNNs to embedded devices is extremely low. Embedded terminal devices must be small and fast to maintain the detection accuracy. Second, with the rapid development of cloud services and IoT, non-information-related industries, such as medical and health information, have become increasingly dependent on information systems and networks, leading to constant risks and threats to information security. Therefore, the protection of information security and the cost of network transmission cannot be ignored. Accordingly, in the early stage of the experiment, two lightweight CNNs (MobileNet V2 and NASNetMobile) were first used for image recognition experiments. After repeated training, it was found that the proper adjustment of the model parameters had almost no effect on the accuracy of the model. Finally, there is still no method to improve the recognition performance

of lightweight CNNs through transfer learning. Considering that the data used in this study are medical images that are highly complex in detail, the nature of the images is different from that of general images.

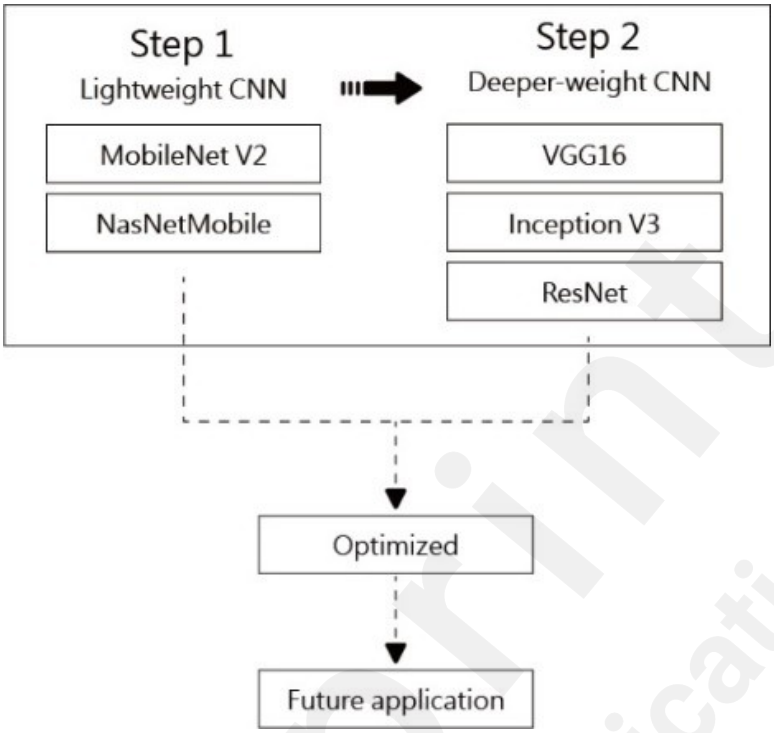


Figure 1. Two-stage experimental process in this study.

In the second stage of the experiment, we decided to use other deeper-weight CNN models for image recognition experiments to improve the nonlinear expression capabilities and learn more complex features. Three deeper-weight CNN models were tested with SPECT images: VGG16, Inception V3, and ResNet. The three types of CNNs were also adjusted for the three influencing factors to optimize their performance. It is expected that the most suitable CNN architecture for SPECT images can be obtained through a comparative analysis. The experimental results can serve as a reference for future research and the detection of AD based on medical image recognition.

Image dataset of SPECT

In general, the collection of image data, which is suitable for developing a brain imaging recognition system, is one of the hardest steps in the procedure. This study used SPECT imaging data with Tc-99m ethyl cysteinate dimer as a tracer to test the effectiveness of the proposed method. The images used in this study were obtained from a single hospital to avoid deviations between different hospital sources. Images were obtained from the archives of the Department of Neurology, National Cheng Kung University Hospital (Figure 2). Brain scan image data were collected from 36 male and 63 female subjects. Experienced neurologists evaluated these image data for the possibility of AD and the detailed diagnosis, including the education level, Mini Mental State Exam (MMSE) score, Cognitive Abilities Screening Instrument (CASI) score, CDR score, and Sum of Box (SOB). The detailed demographic and diagnostic data of the subjects are shown in Table 1.

To reduce unnecessary information in the images and reduce model training errors, we cropped all 99 images to remove excess background information. After the preprocessing operation, we obtained a total of 4,711 images of individual slices (as shown in Figure 3). The

format of each cropped SPECT image was a JPG file with a size of 124 × 120 pixels. Of the 99 patients, 52 were categorized as having questionable dementia (CDR = 0.5), 39 had MCI (CDR = 1), and eight had moderate cognitive impairment (CDR = 2). All 4,711 images were classified into three categories based on the CDR. In this study, 4,461 images were used as training data (80%) and verification data (20%) to perform all the experiments, while the remaining data were manually used as test data (250 images). The data ratio of the training, verification, and test data was approximately 14:3:1.

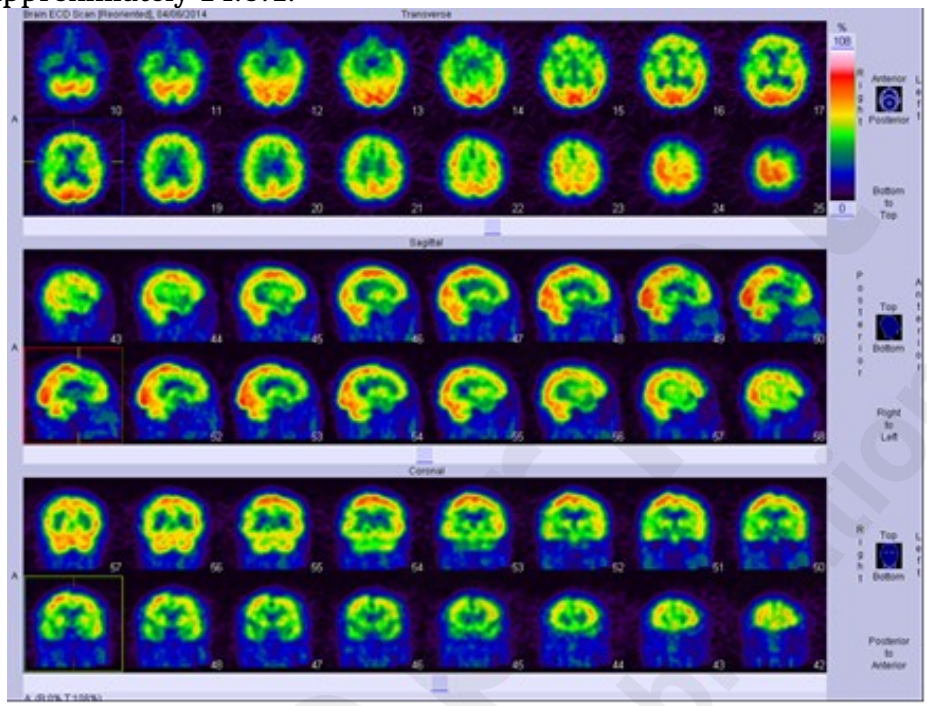


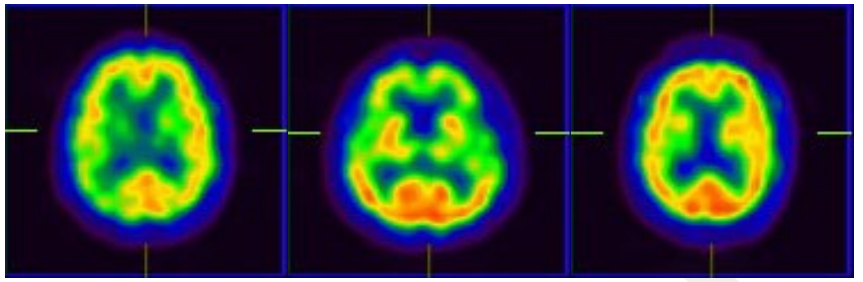
Figure 2. SPECT images.

Table 1. Subjects' demographic and clinical information.

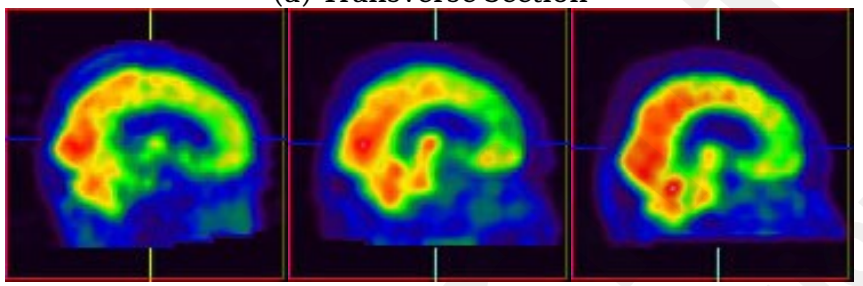
	CDR (0.5)	CDR (1)	CDR (2)	*P* value
N	52	39	8	
Age (years)	72.08 ± 7.96	76.15 ± 8.91	80.11 ± 3.22	0.01*(CDR(0.5)<CDR(2))
Gender (female / male)	27 / 25	31 / 8	5 / 3	0.03*
Education (years)	8.43 ± 4.40	7.03 ± 4.13	7.00 ± 6.08	n.s.
SOB	1.30 ± 0.72	4.73 ± 1.31	10.39 ± 0.93	<0.01**(CDR(0.5)<CDR(1); CDR(1)<CDR(2))
CASI	76.96 ± 11.80	63.26 ± 13.68	48.89 ± 14.91	<0.01**(CDR(0.5)>CDR(1); CDR(1)>CDR(2))
MMSE	22.92 ± 4.04	19.18 ± 3.91	14.78 ± 3.96	<0.01**(CDR(0.5)>CDR(1); CDR(1)>CDR(2))

Values are numbers or mean ± standard deviation (range). Abbreviations: CDR = Clinical Dementia Rating Scale; MMSE = Mini Mental State Exam; N = Number; CASI = Cognitive Abilities

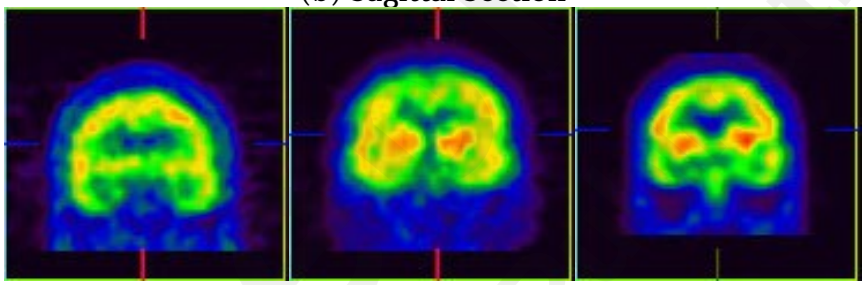
Screening Instrument; SOB = Sum of Box. Data were compared using the chi-squared test and one-way analysis of variance with Bonferroni's post-hoc test. *P < 0.05, **P < 0.01, P > 0.05: no significant difference (n.s.).



(a) Transverse Section



(b) Sagittal Section



(c) Coronal Section

Left: CDR = 0.5; Middle: CDR = 1; Right: CDR = 2

Figure 3. Three different sections of SPECT images.

Architecture of five CNN models after transfer learning

Deep learning frameworks have been applied in medical imaging owing to the success of deep learning in computer vision applications. Unlike general natural image recognition tasks, medical image analysis lacks large labeled training datasets.[38] In particular, the small number of labeled databases is one of the main factors that makes training the classification model difficult because of overfitting problems. Therefore, in the medical field, many transfer learning-based methods have been proposed to address this lack of data.[31] Transfer learning is a process that applies a previously learned model to another field, which helps improve the learning method of few labels or a small number of datasets in the target data. Pre-trained CNNs (OverFeat) have been used in transfer learning to successfully identify and detect vertical pathologies using X-ray and CT modalities.[39] Studies have presented a breast imaging system based on deep CNNs with transfer learning, identifying the specific utility of transfer learning in computer-aided diagnosis.[16] Recently, studies have used a fully convolutional network with transfer learning to perform retinal blood vessel segmentation tasks.[40] When a small dataset is directly trained with deep learning networks, overfitting can easily occur. Transfer learning can improve the initial ability of extracting features to alleviate the risk of overfitting.[31] In this case, transfer learning between task domains is desirable. In addition, fine-tuned CNNs after transfer learning should always be the preferred option, regardless of the size of the available training

sets. In addition, the fine-tuned CNN model after transfer learning can quickly attain the maximum performance.[41] On the contrary, CNNs trained from scratch require extensive training to achieve the highest performance. Therefore, this study used transfer learning to conduct experiments. Such a method can also be executed even with limited training data.

This study used transfer learning to load the weights of the pre-trained model on the new network structure and then trained the network to recognize SPECT images. Five CNNs, i.e., MobileNetV2, NASNetMobile, VGG16, InceptionV3, and ResNet, were used. The experiment modified the structure of the pre-trained model through fine-tuning and was used as the initial model for the SPECT image recognition task. First, we preprocessed the SPECT image data and then used the original convolutional layer in the network structure to extract bottleneck features. Second, we connected the fine-tuned fully connected layer to form a new network structure. The experimental process is illustrated in Figure 4. The fine-tuning process of the fully connected layers in the different CNN models is presented in the following subsections.

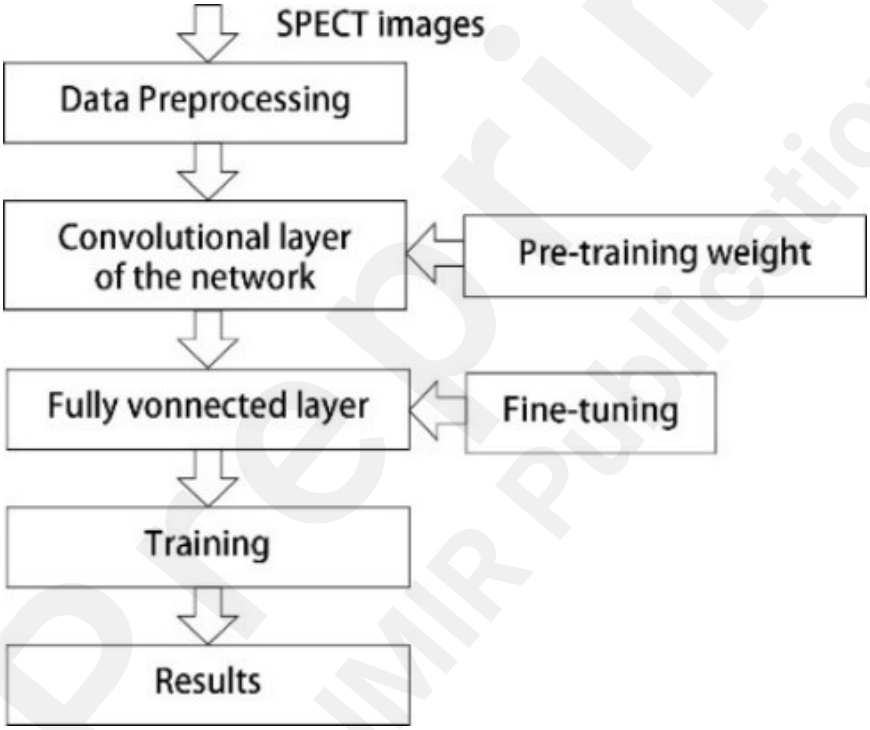


Figure 4. Process and architecture of CNN models.

MobileNetV2

1. Fine-tuning

To prevent excessive model parameters from causing overfitting problems, GlobalAveragePooling2D was added at the end of the model to replace the fully connected layer, and the average value of each feature map was taken as the output. Then, two layers of batch normalization were added to ensure that the input of each layer of the network had the same distribution, which not only reduced the occurrence of overfitting but also accelerated the convergence speed of the training process. The output dimension of DenseNet was set to 4,096. Finally, the dropout layer was added, and the ratio was set to 0.5 to improve the generalization of the model and avoid relying too much on certain regional features. Table 2 lists the layer settings at the end of MobileNetV2 after fine-tuning.

Table 2. Layer settings after fine-tuning and pre-processing for data enhancement.

(a) MobileNetV2		
	Layer	*Type*
Layer Setting	global_average_pooling2d_1	GlobalAveragePooling2D
	batch_normalization_1	BatchNormalization
	dense_1	Dense
	batch_normalization_2	BatchNormalization
	dropout_1	Dropout
Data Enhancement	height_shift_range= 0.2	
	width_shift_range= 0.2	
	shear_range= 0.2	
	zoom_range= 0.2	
	horizontal_flip= True	
	fill_mode= nearest	
(b) NasNetMobile		
	Layer	*Type*
Layer Setting	global_average_pooling2d_1	GlobalAveragePooling2D
	dense_1	Dense
	dropout_1	Dropout
Data Enhancement	height_shift_range= 0.2	
	shear_range= 0.2	
	horizontal_flip= True	
	fill_mode= nearest	
(c) VGG16		
	Layer	*Type*
Layer Setting	flatten	Flatten
	dense_1	Dense
	dropout_1	Dropout
	dense_2	Dense
Data Enhancement	rotation_range= 40	
	height_shift_range= 0.2	
	shear_range= 0.2	
	zoom_range= 0.2	
	horizontal_flip= True	
	fill_mode= nearest	
(d) InceptionV3		
	Layer	*Type*
Layer Setting	global_average_pooling2d_1	GlobalAveragePooling2D
	dense_1	Dense
Data Enhancement	height_shift_range= 0.2	
	horizontal_flip= True	
	fill_mode= nearest	
(e) ResNet		
	Layer	*Type*
Layer Setting	flatten	Flatten
	batch_normalization_1	BatchNormalization
	dense_1	Dense

	dropout_1	Dropout
	dense_2	Dense
Data Enhancement	height_shift_range= 0.2	
	width_shift_range= 0.2	
	horizontal_flip= True	
	fill_mode= nearest	

2. Experimental setup

Because the experimental model used MobileNetV2 as the basic structure, the input image size of MobileNetV2 must be 224×224. Therefore, rescale = 1/255 was added in the preprocessing stage of the image data. Then, each pixel value of the original 124×120 SPECT image was multiplied by a scaling factor to facilitate the convergence of the model. Simultaneously, the dataset was normalized to improve the data integrity. The appearance and reading methods of all image data in all the records were ensured to be identical. To present the classification results in percentages, SoftMax was chosen as the model classifier. The optimization method of the model is Adam to increase the training speed. In the parameter part, the learning rate was set to 10^{-4} 10^{-4}; the exponential decay rates of the first and second orders were 0.9 and 0.999, respectively, by default; epsilon was set to 10^{-8} 10^{-8}; and the attenuation value was set to 10^{-5} 10^{-5} using a low learning rate to retrain the new data. Categorical cross-entropy was chosen as the loss function, batch size was set to 32, and epoch number was set to 50. Before data analysis, insufficient sample data may lead to poor results after training and overfitting problems. To expand the dataset, this study used the data enhancement method to horizontally flip, randomly zoom, and fill the original image (as shown in Table 2) to solve the data imbalance and insufficient data problems.

NASNetMobile

1. Fine-tuning

As described in MobileNetV2, to prevent overfitting caused by excessive model parameters and the limitation of the model input size, GlobalAveragePooling2D was added at the end of the model to replace the final fully connected layer, and the average value of each feature map was used as the output to achieve dimensionality reduction. Simultaneously, the feature information extracted from the previous convolutional and pooling layers was retained. The output dimension of the dense layer was set to 2048. Finally, a dropout layer was added, and the ratio was set to 0.5 to randomly disconnect a certain ratio of the input during the model training process and improve the generalization of the model to avoid relying excessively on certain regional features (as shown in Table 2).

2. Experimental setup

As described in MobileNetV2, rescale = 1/255 was added in the preprocessing stage of the image data to facilitate the convergence of the model. To present the classification results as a percentage, SoftMax was added as the result classifier. Then, stochastic gradient descent (SGD) was selected as the optimizer of the model to increase the training speed. In the parameter part, the learning rate was set to 10^{-4} 10^{-4}, momentum was set to 0.9, and attenuation value was set to 10^{-6} 10^{-6}. Categorical cross-entropy was used as the loss function, batch size was set to 32, and epoch number was set to 50. Before the data analysis, insufficient sample data may lead to poor results after training and overfitting problems. To increase the training data, this study used the data enhancement method to horizontally flip, randomly zoom, and fill the original image (as shown in Table 2) to generate completely different images to solve the imbalance and limited data.

VGG16

1. Fine-tuning

First, a flattened layer was added to dimensionalize the output of the previous convolutional layer into a two-dimensional matrix. This method can reduce the size of an image without losing important image features. The output dimension of DenseNet was set to 1024. Finally, the dropout layer was added, and the ratio was set to 0.5 to improve the generalization of the model and avoid relying excessively on certain regional features (as shown in Table 2).

2. Experimental setup

We also normalized the dataset to improve the integrity of the data and ensure that the appearance and reading methods of all image data in all the records were the same. SoftMax was also used as the resulting classifier. For VGG16, the experimental settings (e.g., Adam optimization method, learning rate, exponential decay rate, and attenuation value) are the same as those of MobileNetV2. The activation function was ReLU. To increase the training data, we performed rotation, shearing strength, horizontal flip, random scaling, filling, and other processes on the original image (as shown in Table 2).

InceptionV3

1. Fine-tuning

To prevent excessive model parameters from causing overfitting problems, GlobalAveragePooling2D was added at the end of the model to replace the fully connected layer, and the average value of each feature map was taken as the output. Then, a dense layer was added, and the output dimension was set to 2048. Please refer to Table 2 for further details.

2. Experimental setup

The input image size of InceptionV3 is 299×299 pixels. Therefore, rescale = 1/255 was added in the preprocessing stage of the image data, and each pixel value of the SPECT image with an original size of 124×120 pixels was multiplied by the scaling factor to facilitate the convergence of the model. To present the classification results as percentages, SoftMax was added as the resulting classifier, and the classification category was set to 3. Similar to the NASNetMobile model, SGD was selected as the model's optimizer to increase the training speed. The learning rate was set to 10^{-5}, momentum was 0.9, loss function was categorical cross-entropy, batch size was 64, and epoch number was 50. Table 2 shows the horizontal flip, random zoom, and original image for data enhancement.

ResNet

1. Fine-tuning

The flattened layer was added to dimensionalize the output of the previous convolutional layer into a two-dimensional matrix, and batch normalization was then added. A dense layer was included, and the output dimension was set to 256. Finally, the dropout layer was also set to 0.5 to avoid relying too much on certain regional features. Table 2 also shows the layers added at the end of the ResNet after fine-tuning. As an illustration, Figure 5 shows the architecture of the ResNet model.

2. Experimental setup

As described in NASNetMobile, rescale = 1/255 was added in the preprocessing stage of the image data to facilitate the convergence of the model. To present the classification results as percentages, SoftMax was chosen as the result classifier, and SGD was chosen as the optimization method. The learning rate was set to 10^{-4}, momentum was 0.9, loss function was categorical cross-entropy, batch size was set to 32, epoch number was set to 50, and activation

function was ReLU. Before the data analysis, we used the data enhancement method to perform position shift, horizontal flip, and fill processing on the original image (as shown in Table 2).

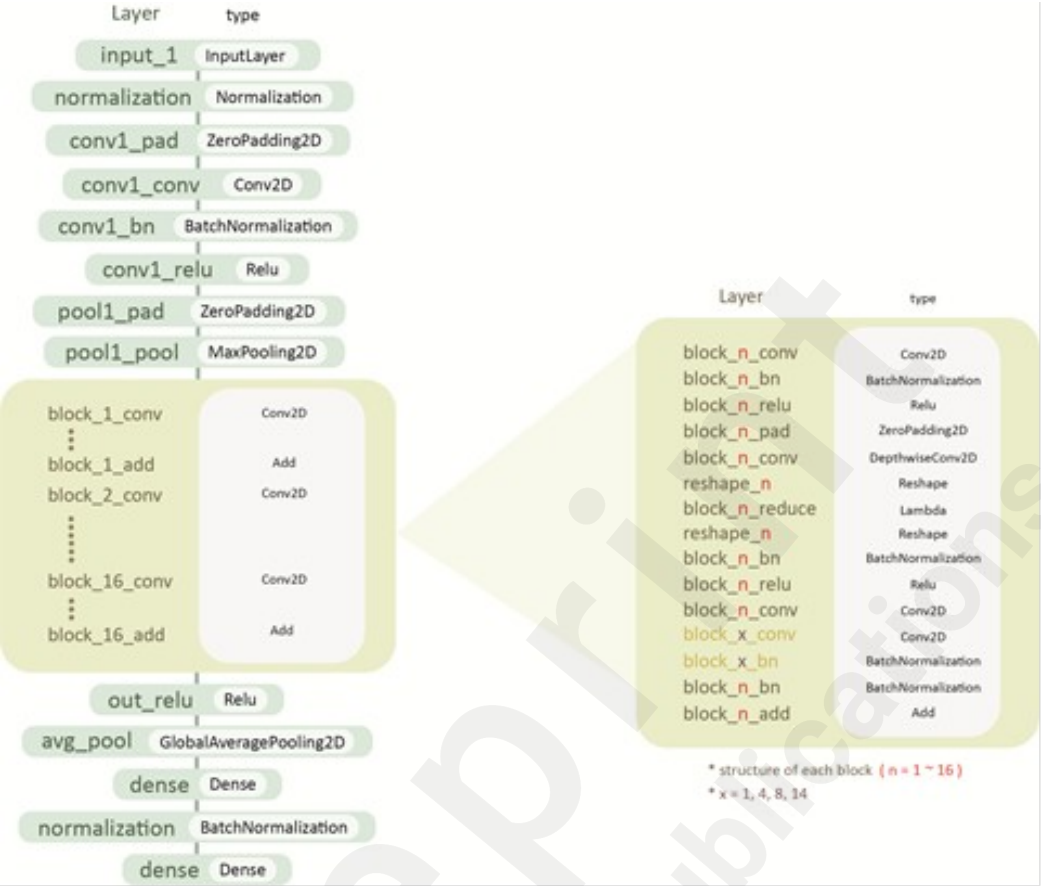


Figure 5. Architecture of the ResNet model.

Results

The demographic characteristics of the study subjects are provided in Table 1. Subjects with CDRs of 2 were significantly older than those with CDRs of 0.5. The SOB scores of subjects with CDR of 2 were significantly higher than those of subjects with CDRs of 1 and 0.5. The CASI and MMSE scores of subjects with CDR of 2 were significantly lower than those of subjects with CDRs of 1 and 0.5 (Table 1). The scores of the SOB, CASI, and MMSE validated the severity of cognitive impairments classified using the CDR score. Following the evaluation of the severity of cognitive impairments, SPECT images were analyzed in this study using two lightweight and three deeper-weight CNN models to distinguish the severity of AD in patients based on the CDR scores. We used two evaluation indicators, i.e., accuracy and loss, to evaluate the detection performance of the model. This study first examined the impact of different sections of brain images (i.e., the transverse, sagittal, and coronal sections) on the model to improve its performance in detecting the severity of AD by detecting different sections of SPECT images. There were a total of 1602 transverse images, 1584 sagittal images, and 1525 coronal images. Subsequently, the SPECT image data of the three different cross-sections were mixed for the model identification experiments. Table 3 shows the performance indicators of the five CNN models that identify different CDR scores from a single brain section and a mixed section on the validation and testing data. Based on the classification results for a single brain section shown in Table 3, the ResNet model was the best performer among the five different CNN models. The detection accuracy rates in the transverse, sagittal, and coronal section image data were 61%, 63%, and 65%, respectively. InceptionV3 is the worst-performing CNN model, with detection

accuracy rates of 54%, 51%, and 48%. The classification accuracy of MobileNetV2, NASNetMobile, and VGG16 was approximately 60%.

Table 3. Detection performance of two lightweight and three deeper-weight CNN models

Model	Types of data	Val_acc (%)	Loss function (Val)	Test_acc (%)	Loss function (Test)
MobileNet V2	Transverse section	60.57	0.1952	60.01*	0.2013
	Sagittal section	58.69	0.3368	50.45	0.4567
	Coronal section	60.89	0.1876	57.77	0.1935
	Mixed (three kinds of section)	61.87*	0.6501	58.43	0.1268
NASNetMobile	Transverse section	63.87*	0.6301	61.25*	0.4322
	Sagittal section	57.12	0.2615	57.03	0.1623
	Coronal section	63.03	0.1246	55.49	0.2761
	Mixed (three kinds of section)	59.89	0.3321	58.77	0.2311
VGG16	Transverse section	66.03	0.2388	64.58	0.2564
	Sagittal section	64.20	0.4601	61.89	0.1733
	Coronal section	58.66	0.4245	53.01	0.3014
	Mixed (three kinds of section)	69.45*	0.1003	67.53*	0.1318
InceptionV3	Transverse section	56.77*	0.1388	54.22*	0.2108
	Sagittal section	53.09	0.3350	50.98	0.2861
	Coronal section	52.78	0.2018	47.55	0.1937
	Mixed (three kinds of section)	54.03	0.4003	52.13	0.4766
ResNet	Transverse section	67.23	0.2387	61.20	0.3566
	Sagittal section	65.37	0.1564	63.37	0.1804
	Coronal section	68.51	0.1956	65.28	0.2390
	Mixed (three kinds of section)	72.39**	0.1186	70.79**	0.2130

For the classification results of the mixed section, Table 3 shows that the best model is ResNet, with a test accuracy rate of 71%. Although the accuracy rate was not very high (e.g., over 80% or more), it was sufficient to judge the performance of different CNN models for medical images. The second-highest accuracy rate was 68% (VGG16). InceptionV3 was the worst-performing model, with a detection accuracy rate of only 52%. Based on the experimental results, the lightweight CNN models, i.e., MobileNetV2 and NASNetMobile, were not

satisfactory. For the image data obtained by mixing the three sections, the best validation accuracies of MobileNetV2 and NASNetMobile were only 58% and 59%, respectively, indicating that there is much room for improvement for the two lightweight CNN models to match the SPECT medical images.

Discussion

In recent years, neuroimaging has become a beneficial method for understanding the pathogenesis of AD, and the use of deep learning techniques in the field of medical imaging technologies has made progress in research and clinical care. Many researchers have been engaged in the development of auxiliary diagnostic systems for medical imaging. In this study, five different CNN models were trained in the transfer learning process, and performance evaluations and comparisons were performed. The main goal of this study is to compare the detection performance of different CNN structures for medical images, which was confirmed based on the results. Moreover, compared with using a single cross-sectional image as the input, the use of mixed data from the three cross-sections as the model input produced better results. The validation and test accuracy records in Table 3 show that the performance of ResNet is better than those of MobileNet V2, NASNetMobile, VGG16, and Inception V3. In this type of medical imaging dataset, the deeper-weight network performed better than the lightweight network. In addition, ResNet can effectively avoid gradient explosion, disappearance, and network degradation. Therefore, ResNet can be used to develop an AI expert system that can analyze the severity of AD using SPECT images, as shown in Figure 6. The AI expert system does not replace physicians but helps them make clearer decisions on disease classification and more confident diagnosis based on systematic objective information to reduce the uncertainty in disease classification diagnosis.

(Physicians can import the SPECT image (left part), and the AI expert system will show the outcome of 0.5 (right part) to help them make decisions on disease classification.)

Figure 6. AI expert system based on the ResNet model.

The performance of CNN models, such as the ResNet model, varied depending on the type of study, field of study, data used, and imbalance in the sample.[42] In neuroimaging research, there have been many applications in image processing and feature recognition applied to the classification of AD. Among the various deep-learning methods, ResNet has been widely used for the classification and diagnosis of AD. Amin-Naji et al.[43] used a residual structure in each branch of a CNN. The OASIS dataset was used to evaluate the effectiveness of the model. Finally, an accuracy of 98.72% was obtained in the classification of old patients with AD and normal controls using MRI, which is the best result compared with those in other previous studies on the same database. Abrol et al.[44] tested the ability of a residual structure to learn from structural MRI data using neuroimaging, showing whether it is binary classification or heap classification. Their results showed that the recognition performance of the ResNet architecture is better than that of the SVM and SAE methods. Karasawa and Ohwada[45] proposed a ResNet-based structure for classifying MRI data from the Alzheimer's Disease Neuroimaging Initiative database. The experimental results show that the proposed 39-layer residual structure has the highest accuracy compared to those of VGGNet and ResNet-50. Therefore, related research has shown that the residual structure has a certain effect on the applicability of image recognition in neuroimaging. This finding shows that the ResNet structure proposed in this study is effective for AD classification. In previous studies, the performance of the total and subscale scores of the Mattis Dementia Rating Scale for discriminating MCI from NC, MCI from mild AD, and mild AD from moderate AD revealed a sensitivity and specificity ranging from 40% to 87%[46], which were comparable with the performance of our rCBF SPECT CNN model.

This study selected image data with different categories but similar characteristics for the analysis. In the data collection process, there are few data categories with CDR = 3, so this category of data was excluded when training the model to avoid affecting the performance of the CNN model. Therefore, in future research, the image data of this category will be considered to complete all the data categories and analyze their impact on the performance of the CNN model. Experimental research on different brain cross-sectional images shows that the classification accuracy obtained by training five CNN models with a single cross-sectional image is no better than that obtained using mixed data. Moreover, there is a significant gap in the results of generating three types of slices as the dataset at the same time. The main reason for this difference may be the variation in the amount of data used. In this study, the best model with a test accuracy rate of 71% appears to be insufficient (higher), which can be summarized as follows:

1. This is the first time that these five commonly used CNN models have been tested for transfer learning in identifying MCI, mild AD, and moderate AD scores based on SPECT images.
2. The medical image data used in this study are not similar to the dataset of the pre-training model; therefore, the feature extraction of the pre-training model is different, which may lead to poor training results.
3. The SPECT image data used in this study contained 4711 images, and few images may have led to poor training results.
4. The number of layers and the depth of the pre-training model used in this study were not very large. Therefore, insufficient network depth and computing resources may also lead to poor training results.

Conclusion

The purpose of this study is to explore different CNN models for classifying AD severity based

on SPECT images, which have a lower cost than CT and MRI. This study used deep and transfer learning to test the performance of two lightweight and three deeper-weight CNN models for SPECT image data classification, showing that ResNet is the most suitable CNN model. The results indicate that the performance of deeper CNN models is better than that of lightweight CNN models. Although this study has proven that deep learning technology can achieve a good performance in SPECT image classification, there is still room for further improvement. The method proposed in this study provides important implications for image recognition and deep learning in the development of mobile applications in AI and medical treatment and has great potential in other biomedical fields. This deep learning method not only opens up new ways for medical image analysis but also enables researchers and physicians to predict any new data in the future. In the future, we hope to collect more clinical data from different hospitals to increase the depth of the training dataset. Moreover, neuroimaging information can be combined with functional information of the cognitive scale to obtain a better machine learning model for the classification of AD severity.

Acknowledgments

This research was supported by the Ministry of Science and Technology, Taiwan, under Grant MOST108-2628-E-006-MY3 for the 3-year project, MOST110-2314-B-006-096- for the 1-year project, and MOST111-2314-B-006 -064 -MY3 for the 3-year project.

Conflicts of Interest

The author reports no conflicts of interest in this work.

Ethics declarations

This study was approved by the National Cheng Kung University Human Research Ethics Committee (NCKU HREC-E-108-282-2) and conformed to the principles of the Declaration of Helsinki. All participants provided written informed consent prior to participating in the study.

Abbreviations

AD: Alzheimer’s disease
AI: artificial intelligence
CASI: Cognitive Abilities Screening Instrument
CDR: Clinical Dementia Rating
CNNs: convolutional neural networks
CT: computer-assisted tomography
IoT: Internet of things
MCI: mild cognitive impairment
MMSE: Mini Mental State Exam
MRI: magnetic resonance imaging
NNs: Neural networks
rCBF: regional cerebral blood flow
SOB: Sum of Box
SPECT: single-photon emission computed tomography

References

1. Das, T. K.; Kumar, P. M. Big data analytics: A framework for unstructured data analysis. Int. J. Eng. Sci. 2013, 5: 153.
2. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. Shufflenet: An extremely efficient convolutional neural network for mobile devices. Proceedings of the IEEE Conference on Computer Vision

and Pattern Recognition; 2018: 6848–6856.
3. Pan, S. T.; Lai, C. C. Identification of chaotic systems by neural network with hybrid learning algorithm. Chaos, Solitons and Fractals. 2008, 37: 233–244.
4. Yao, X. Evolving artificial neural networks. Proceedings of the IEEE; 1999, 87(9): 1423–1447.
5. Feng, S.; Zhou, H.; Dong, H.B. Using deep neural network with small dataset to predict material defects. Mater. Des. 2018: 162. 10.1016/j.matdes.2018.11.060.
6. Lin, Y.-C.; Lai, H.-H.; Yeh, C.-H. Neural network models for product image design. Lect. Notes Comput. Sci. 2004, 3215: 618–624.
7. Wei, C.-C.; Ma, M.-Y.; Lin, Y.-C. Applying Kansei engineering to decision making in fragrance form design. Smart Innov. Syst. Technol.; 2011, 10: 85–94.
8. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; ... & Berg, A.C. Imagenet large scale visual recognition challenge. Int. J. Comput. Vis. 2015, 115: 211–252.
9. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. Nature. 2015, 521: 436–444.
10. LeCun, Y.; Boser, B.E.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.E.,; Jackel, L.D. Handwritten digit recognition with a back-propagation network. Advances in Neural Information Processing Systems; 1990: 396–404.
11. Taner, A.; Öztekin, Y.B.; Duran, H. Performance analysis of deep learning CNN models for variety classification in hazelnut. Sustainability. 2021, 13: 6527. https://doi.org/10.3390/su13126527
12. Wang, P.; Xiao, J.; Kawaguchi, K.; Wang, L. Automatic ceiling damage detection in large-span structures based on computer vision and deep learning. Sustainability. 2022, 14: 3275. https://doi.org/10.3390/su14063275
13. Hertz, J.A. Introduction to the theory of neural computation. CRC Press, 2018.
14. Chen, T.; Chen, Y.; Lv, M.; He, G.; Zhu, T.; Wang, T.; Weng, Z. A payload based malicious HTTP traffic detection method using transfer semi-supervised learning. Appl. Sci. 2021, 11: 7188. https:// doi.org/10.3390/app11167188
15. Gu, J.; Na, J.; Park, J.; Kim, H. Predicting success of outbound telemarketing in insurance policy loans using an explainable multiple-filter convolutional neural network. Appl. Sci. 2021, 11: 7147. https://doi.org/10.3390/app11157147
16. Sun, Y.; Ma, S.; Sun, S.; Liu, P.; Zhang, L.; Ouyang, J.; Ni, X. Partial discharge pattern recognition of transformers based on MobileNets convolutional neural network. Appl. Sci. 2021, 11: 6984. https:// doi.org/10.3390/app11156984
17. Dhiman, G.; Juneja, S.; Viriyasitavat, W.; Mohafez, H.; Hadizadeh, M.; Islam, M.A.; El Bayoumy, I.; Gulati, K. A novel machine-learning-based hybrid CNN model for tumor identification in medical image processing. Sustainability 2022, 14: 1447. https://doi.org/10.3390/su14031447
18. Hill, D.L.; Batchelor, P.G.; Holden, M.; Hawkes, D.J. Medical image registration. Phys. Med. Biol. 2001, 46: R1.
19. Larner, A.J. Getting it wrong: The clinical misdiagnosis of Alzheimer's disease. Int. J. Clin. Pract. 2004, 58: 1092–1094.
20. Aël Chetelat, G.; Baron, J.C. Early diagnosis of Alzheimer's disease: Contribution of structural neuroimaging. Neuroimage. 2003, 18: 525–541.
21. Hashizume, Y. Macroscopic findings of brain with dementia. Neuropathology. 2022. **n/a**(n/a).
22. Li, F., et al., A robust deep model for improved classification of AD/MCI patients. IEEE J. Biomed. Health Inform. 2015, 19(5): 1610–1616.
23. Leung, G.M., et al., The Economics of Alzheimer Disease. Dement Geriatr Cogn Disord. 2003, 15(1): 34–43.

24. Binnewijzend, M.A.; Benedictus, M.R.; Kuijer, J.P.; van der Flier, W.M.; Teunissen, C.E.; Prins, N.D.; ... Barkhof, F. Cerebral perfusion in the predementia stages of Alzheimer's disease. Eur. Radiol. 2016, 26: 506–514.
25. Higdon, R.; Foster, N.L.; Koeppe, R.A.; DeCarli, C.S.; Jagust, W.J.; Clark, C.M.; ... Minoshima, S. A comparison of classification methods for differentiating fronto-temporal dementia from Alzheimer's disease using FDG-PET imaging. Stat Med. 2004, 23: 315–326.
26. Bonte FJ, Weiner MF, Bigio EH, White CL, 3rd. Brain blood flow in the dementias: SPECT with histopathologic correlation in 54 patients. Radiology. 1997, 202(3): 793–797.
27. Arnáiz, E.; Almkvist, O. Neuropsychological features of mild cognitive impairment and preclinical Alzheimer's disease. Acta Neurol. Scand. 2003, 107: 34–41.
28. Huang, H.-C., et al., Diagnostic accuracy of the Clinical Dementia Rating scale for detecting mild cognitive impairment and dementia: A bivariate meta-analysis. Int. J. Geriatr. Psychiatry. 2021. 36(2): 239–251.
29. Tobore, I., et al., Deep learning intervention for health care challenges: Some biomedical domain considerations. JMIR Mhealth Uhealth. 2019. 7(8): e11966.
30. Huynh, B.Q.; Li, H.; Giger, M.L. Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. J. Med. Imaging. 2016, 3: 034501.
31. Raghu, M.; Zhang, C.; Kleinberg, J.; Bengio, S. Transfusion: Understanding transfer learning for medical imaging. Adv Neural Inf Process Syst. 2019: 3347–3357.
32. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. Mobilenetv2: Inverted residuals and linear bottlenecks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2018: 4510–4520.
33. Zoph, B.; Vasudevan, V.; Shlens, J.; Le, Q.V. Learning transferable architectures for scalable image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2018: 8697–8710.
34. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv 2014, arXiv:1409.1556.
35. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016: 2818–2826.
36. Targ, S.; Almeida, D.; Lyman, K. ResNet in ResNet: Generalizing residual architectures. arXiv 2016, arXiv:1603.08029.
37. Yu, X., et al., Transfer learning for medical images analyses: A survey. Neurocomputing, 2022. 489: 230–254.
38. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; ... Sánchez, C.I. A survey on deep learning in medical image analysis. Medical Image Analysis; 2017, 42: 60–88.
39. Van Ginneken, B.; Setio, A.A.; Jacobs, C.; Ciompi, F. Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans. 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI) 2015; 286–289.
40. Jiang, Z.; Zhang, H.; Wang, Y.; Ko, S.B. Retinal blood vessel segmentation using fully convolutional network with transfer learning. Comput Med Imaging Graph. 2018, 68: 1–15.
41. Shin, H.C.; Roth, H.R.; Gao, M.; Lu, L.; Xu, Z.; Nogues, I.; ... Summers, R.M. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans Med Imaging. 2016, 35: 1285–1298.
42. Johnson, J. and Khoshgoftaar, T. Survey on deep learning with class imbalance. J. Big Data. 2019. 6: 27.
43. Amin-Naji, M.; Mahdavinataj, H.; Aghagolzadeh, A. Alzheimer's disease diagnosis from

structural MRI using Siamese convolutional neural network. 2019 4th International Conference on Pattern Recognition and Image Analysis (IPRIA); 2019: 75–79.
44. Abrol, A.; Bhattarai, M.; Fedorov, A.; Du, Y.; Plis, S.; Calhoun, V. Alzheimer's disease neuroimaging initiative. Deep residual learning for neuroimaging: An application to predict progression to Alzheimer's disease. J. Neurosci. Methods. 2020: 108701.
45. Karasawa, H.; Liu, C.L.; Ohwada, H. Deep 3D convolutional neural network architectures for Alzheimer's disease diagnosis. Int. J. Intell. Inf. Database Syst. 2018: 287–296.
46. Qian S, Chen K, Guan Q, Guo Q. Performance of Mattis dementia rating scale-Chinese version in patients with mild cognitive impairment and Alzheimer's disease. BMC Neurol. 2021, 21: 172.