

# Screening of Alzheimer's disease by facial complexion using artificial intelligence

Yumi Umeda-Kameyama<sup>1</sup>, Masashi Kameyama<sup>2</sup>, Tomoki Tanaka<sup>3</sup>, Bo-Kyung Son<sup>3,6</sup>, Taro Kojima<sup>1</sup>, Makoto Fukasawa<sup>4</sup>, Tomomichi Iizuka<sup>5</sup>, Sumito Ogawa<sup>1</sup>, Katsuya Iijima<sup>3,6</sup>, Masahiro Akishita<sup>1</sup>

<sup>1</sup>Department of Geriatric Medicine, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

<sup>2</sup>Department of Diagnostic Radiology, Tokyo Metropolitan Geriatric Hospital and Institute of Gerontology, Tokyo, Japan

<sup>3</sup>Institute of Gerontology, The University of Tokyo, Tokyo, Japan

<sup>4</sup>Department of Nuclear Medicine, Fukujuji Hospital, Japan Anti-Tuberculosis Association, Kiyose, Japan

<sup>5</sup>Center for Dementia, Fukujuji Hospital, Japan Anti-Tuberculosis Association, Kiyose, Japan

<sup>6</sup>Institute for Future Initiatives, The University of Tokyo, Tokyo, Japan

**Correspondence to:** Masashi Kameyama; **email:** [kame-tky@umin.ac.jp](mailto:kame-tky@umin.ac.jp)

**Keywords:** artificial intelligence, face, dementia, machine learning

**Received:** November 3, 2020

**Accepted:** December 10, 2020

**Published:** January 25, 2021

**Copyright:** © 2021 Kameyama et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

Despite the increasing incidence and high morbidity associated with dementia, a simple, non-invasive, and inexpensive method of screening for dementia is yet to be discovered. This study aimed to examine whether artificial intelligence (AI) could distinguish between the faces of people with cognitive impairment and those without dementia.

121 patients with cognitive impairment and 117 cognitively sound participants were recruited for the study. 5 deep learning models with 2 optimizers were tested. The binary differentiation of dementia / non-dementia facial image was expressed as a "Face AI score".

Xception with Adam was the model that showed the best performance. Overall sensitivity, specificity, and accuracy by the Xception AI system and AUC of the ROC curve were 87.31%, 94.57%, 92.56%, and 0.9717, respectively. Close and significant correlations were found between Face AI score and MMSE ( $r = -0.599$ ,  $p < 0.0001$ ). Significant correlation between Face AI score and chronological age was also found ( $r = 0.321$ ,  $p < 0.0001$ ). However, MMSE score showed significantly stronger correlation with Face AI score than chronological age ( $p < 0.0001$ ).

The study showed that deep learning programs such as Xception have the ability to differentiate the faces of patients with mild dementia from that of patients without dementia, paving the way for future studies into the development of a facial biomarker for dementia.

## INTRODUCTION

Dementia is one of the most serious problems facing a global aging population. Diagnosis of dementia is important for early intervention of the disease.

However, many diagnostic methods are invasive or are time-consuming. For example, psychological assessment takes time, examination of cerebrospinal fluid is invasive and amyloid positron emission tomography is costly. Thus, there is demand for a

simple, noninvasive, and inexpensive method for screening for dementia.

The perceived age of older adults was shown to be a robust biomarker of aging that is predictive of survival, telomere length [1], DNA methylation [2], carotid atherosclerosis [3], and bone status [4]. It was demonstrated that perceived age reflects cognitive function more closely than chronological age [5]. Furthermore, in the authors' experience, advanced Alzheimer's disease (AD) patients display a specific complexion. Thus, it was postulated that cognitive decline may be expressed in a patient's face.

Deep learning, a major branch of machine learning in artificial intelligence (AI), has remarkably improved the performance and detection capabilities of AI programs since convolutional neural network (CNN) was first developed [6]. The authors have previously reported how CNN could be used to discern AD and dementia with Lewy bodies (DLB) perfusion single photon emission tomography (SPECT) images. The differentiation was made using cingulate island sign [7]. Based on this information, it was hypothesized that AI software may be able to classify patients as having cognitive impairment or not using facial recognition.

The present study aimed to examine whether AI can distinguish facial traits of cognitive impairment patients from that of non-dementia patients. The findings of this study lay the foundations for the development of a non-invasive, inexpensive and rapid screening tool for cognitive impairment using AI.

## RESULTS

### Demographics

3 patients from the Department of Geriatric Medicine were diagnosed as cognitively normal after several tests and were thus included in the non-dementia arm of the study. Although the participants from the Kashiwa cohort live self-sufficiently in the community, one participant showed declining MMSE over the past few years and was thus classified into the cognitive impairment arm of the study. Two patients were diagnosed to be dementia with Lewy bodies (DLB), two were diagnosed to idiopathic normal pressure hydrocephalus (iNPH), one had aphasia due to cerebral infarction. The rest of the participants were diagnosed to be Alzheimer's disease. The five non-AD patients were excluded from the analyses.

Table 1 summarizes the demographics of participants.

### Models examined

Learning curves for models, Xception, SENet50, ResNet50, VGG16, and simple CNN with SGD and Adam optimizer are shown in Figure 1. Xception with Adam showed the best performance with a peak validation accuracy of more than 94% and a bottom validation loss of less than 0.21. Xception with Adam optimizer showed slightly better performance, therefore Adam appeared to be the preferred optimizer for Xception. ResNet, SENet showed fast learning, however validation cross entropy loss did not drop adequately. VGG16 with Adam showed good accuracy, however cross entropy loss could not be calculated in an early epoch. VGG16 with SGD showed successful learning with the peak validation accuracy being approximately 91%. Simple CNN with Adam also showed learning with the peak validation accuracy being approximately 92%. However, validation cross entropy loss of simple CNN did not show stability.

From these results, Xception with Adam applied as an optimizer was chosen for use in this study. 28 epochs was chosen as this was where the bottom of loss was seen in the model.

### Evaluation of the AI model

Sensitivity, specificity, and accuracy by the AI system (boundary: Face AI score = 0) and area under the curve (AUC) of receiver operating characteristic (ROC) curve of the 10 groups were  $87.52 \pm 11.91\%$ ,  $94.57 \pm 10.88\%$ ,  $92.56 \pm 8.22\%$ , and  $0.9827 \pm 0.0201$ , respectively ( $\pm$  denotes standard deviation). The best sensitivity and specificity according to ROC analysis were  $97.80 \pm 3.36\%$ ,  $96.00 \pm 3.43\%$ , respectively.

Overall sensitivity, specificity, and accuracy of the Xception AI system (boundary: Face AI score = 0) and AUC of ROC curve were 87.31%, 94.57%, 92.56%, and 0.9717, respectively (Table 2). The ROC curve is displayed in Figure 2A. Sensitivity and specificity according to ROC analysis were 96.27%, 91.42% (boundary: Face AI score = -1.51).

Median age of the participants was 76. Sensitivity, specificity, accuracy and AUC of ROC curve for the aged ( $>76$ ) were 83.33%, 94.89%, 89.96%, and 0.9687, respectively; those for the relatively young ( $\leq 76$ ) were 100%, 94.37%, 95.10%, 0.9805, respectively (Figure 2B) (Table 2).

### Correlation between face AI score and MMSE/age

Significant close correlations were found between Face AI score and MMSE ( $r = -0.599$ ,  $t = -16.40$ ,  $p = 2.47 \times 10^{-48}$ ;

**Table 1. Demographic features of study participants.**

	Cognitive impairment		Normal		<i>p</i>
	Male	Female	Male	Female	
Participants		121		117	
Sex	46	75	50	67	0.509 <sup>a</sup>
Age	80.6±6.5	81.5±6.2	75.8±6.0	75.7±5.1	1.20 × 10 <sup>-9b</sup>
MMSE score	21.9±5.0	21.3±5.1	28.8±1.8	29.0±1.3	4.23 × 10 <sup>-34b</sup>
photographs	54	80	155	195	
hypertension (+/-)	26/19	57/19	31/19	38/29	0.139 <sup>a</sup>
hyperlipidemia (+/-)	13/33	30/46	25/24	37/30	0.00603 <sup>a</sup>
diabetes (+/-)	13/33	15/61	20/30	25/42	0.0114 <sup>a</sup>
osteoporosis (+/-)	9/37	33/43	19/31	25/33	0.112 <sup>a</sup>

Data are shown as numbers or means ± standard deviation. MMSE, mini-mental state examination; a, Fisher's exact test comparing cognitive impairment and normal; b, one-way analysis of variance.

Figure 3A). Significant correlation between Face AI score and chronological age was also found ( $r = 0.321$ ,  $t = 7.44$ ,  $p = 4.57 \times 10^{-13}$ ; Figure 3B). However, MMSE score showed a significantly stronger correlation with Face AI score than chronological age ( $t = 13.45$ ,  $p = 3.25 \times 10^{-35}$ ; Steiger's test [8]).

The correlation coefficient between MMSE and Face AI score in female ( $r = -0.661$ ) was significantly stronger than that in male ( $r = -0.501$ ) ( $p = 0.00833$ ) when evaluated using Fisher's Z-transformation method [9].

The difference in the correlation coefficient between age and Face AI score by sex (male:  $r = 0.229$ , female:  $r = 0.388$ ) when assessed using the same method was not significant ( $p = 0.0565$ ).

### Analysis of upper and lower half faces

44 epochs was chosen for upper half and 41 epochs for lower half, as these were where the bottom of loss was seen in the model.

Sensitivity, specificity, accuracy and AUC of ROC curve for upper half faces were 85.07%, 93.14%, 90.91%, and 0.9635, respectively; those for the lower half faces were 91.04%, 95.14%, 94.01%, 0.9803, respectively (Figure 2C) (Table 2).

## DISCUSSION

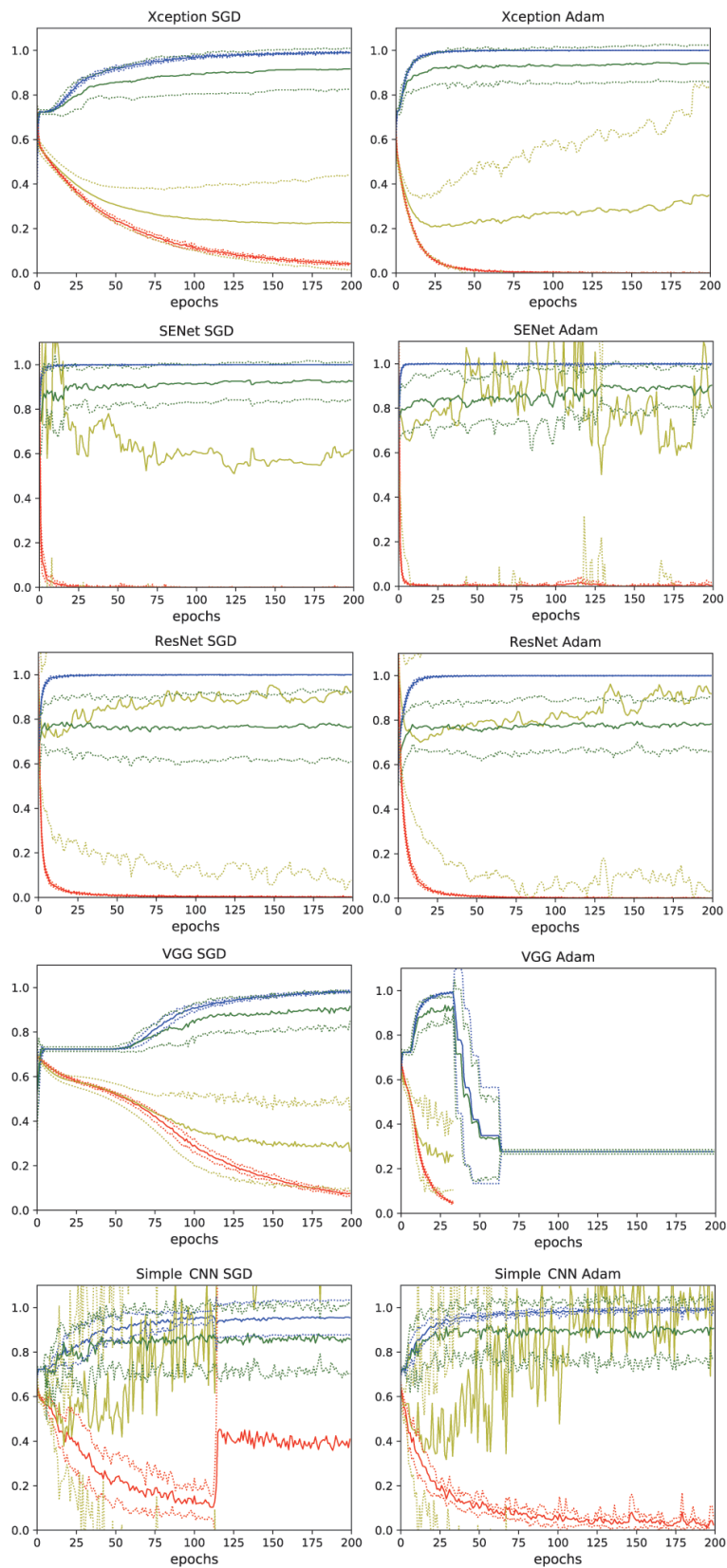
The deep learning network implemented in this study successfully discerned faces of participants with cognitive impairment from those without dementia. With an accuracy of 92.56% and AUC of ROC being 0.9717, this method is reliable enough for implementation as an initial screening test for dementia. The relatively low sensitivity (87.31%), which may be

attributable to the fewer number of dementia participant facial images, could be improved to 96.27% by employing ROC analysis. Face AI score correlated significantly more strongly with MMSE than with age. The weaker correlation between Face AI score and age ( $r = 0.321$ ) would imply that the significant difference between age of dementia participants with and without dementia is unlikely to have affected the AI system. If the AI system relied on participants age, limiting the age range would worsen the results. However, limiting the age did not impair the accuracy (Figure 2B).

The AI system is too complicated and has so-called "black box" nature. Although both upper and lower half faces showed excellent performance, lower half showed slightly better performance than upper half contrary to our expectation (Figure 2C). The system may get more information from mouths or wrinkles than eyes or hair.

The correlation coefficient between Face AI score and MMSE was significantly higher in females than in males. A previous study by the authors demonstrated a significantly stronger correlation between MMSE and perceived age than with chronological age but only in female participants [5], which may be attributable to the tendency for cognitively healthy older women to pay more attention to their appearances relative to their cognitively impaired counterparts, resulting in a marked difference in perceived age.

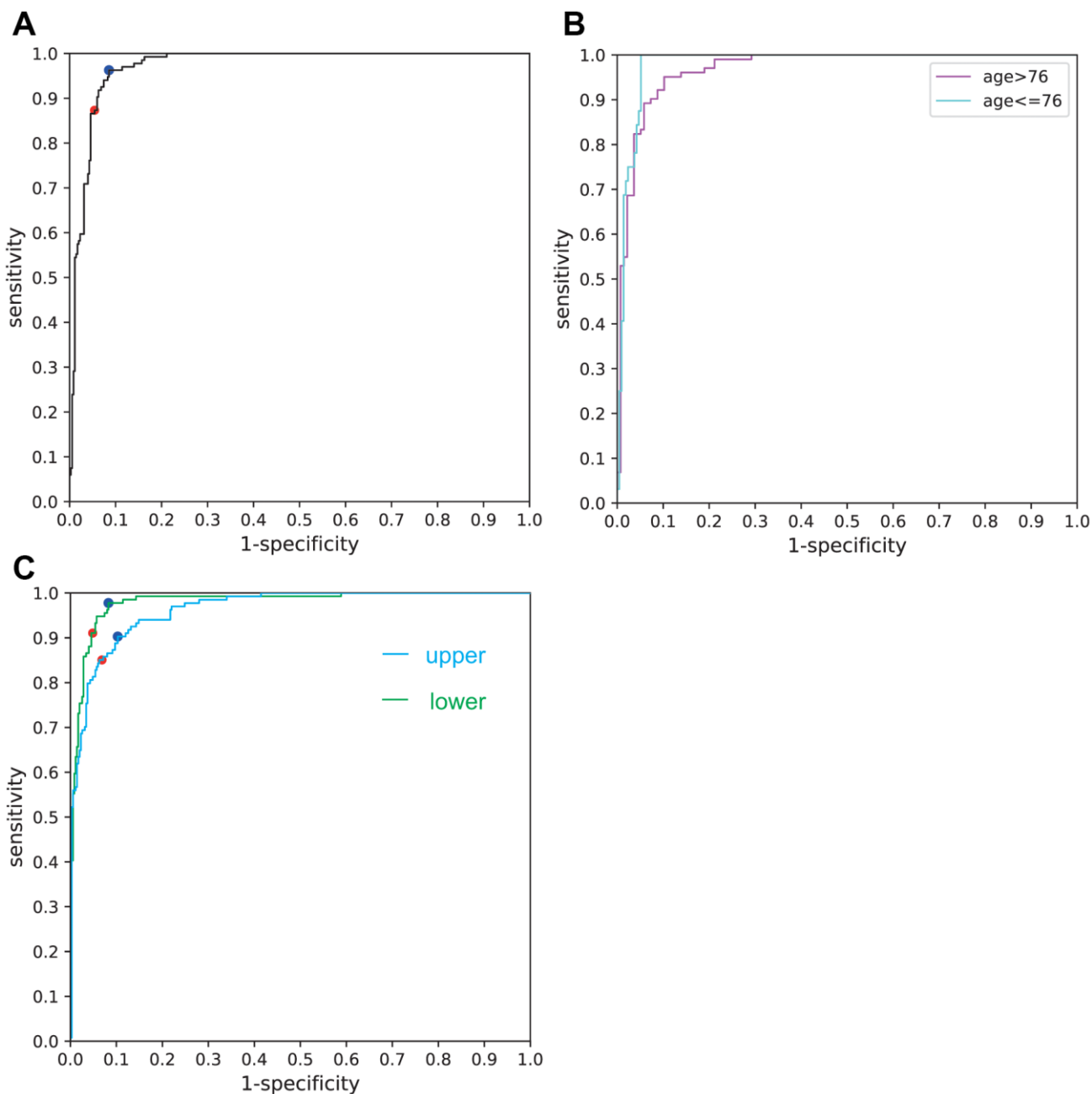
The fact that SENet50, ResNet50, and simple CNN models were unable to reduce loss during discriminating facial images of dementia and non-dementia participants suggests that the task is more difficult for AI software than discerning differences in SPECT images [7]. This is understandable given that while SPECT images can be classified by trained nuclear medicine physicians, diagnosis of dementia



**Figure 1. Learning curves of the deep learning models.** Xception, SENet50, ResNet50, VGG16, and simple CNN with SGD and Adam optimizers were tested. The thin lines denote mean  $\pm$  standard deviation of 10 groups. *blue*: accuracy, *red*: loss, *green*: validation accuracy, *yellow*: validation loss.

**Table 2. Performance of AI.**

	Sensitivity	Specificity	Accuracy	AUC	Best sensitivity	Best specificity	Threshold
total	0.8731	0.9457	0.9256	0.9717	0.9627	0.9143	-1.51
aged (>76)	0.8333	0.9489	0.8996	0.9687	0.9510	0.8978	-1.51
young ( $\leq 76$ )	1	0.9437	0.9510	0.9805	1	0.9484	0.04
upper face	0.8507	0.9314	0.9091	0.9635	0.9030	0.8971	-0.65
lower face	0.9104	0.9514	0.9401	0.9803	0.9776	0.9171	-1.2



**Figure 2.** (A) ROC curve for the overall facial images used. The red dot denotes the sensitivity and specificity calculated by the AI software (boundary: Face AI score = 0). The blue dot denotes the sensitivity and specificity calculated by ROC analysis (boundary: Face AI score = -1.51). (B) ROC curve for the aged (>76) and the relatively young ( $\leq 76$ ). (C) ROC curve for upper half and lower half faces.

from facial images alone cannot be done manually. Suitable deep learning networks such as Xception are required to accomplish the complex task of detection of dementia through facial images.

This study has several limitations. Firstly, the study comprised only 484 images as it was performed at a single institution with a single cohort. Moreover, this study may be affected by institutional biases. Although pretraining assisted AI system learning with limited images, further studies employing a larger number of images from multicenter will be needed to confirm the results. Secondly, the non-dementia participants of this study may have had undetected dementia that did not require nursing care. Similarly, though the majority of dementia participants were assumed to have Alzheimer's disease, this was not confirmed using pathology or

amyloid positron emission tomography, and one participant from Kashiwa cohort did not undergo sufficient tests. Thus, some participants may have been suffering other forms of dementia such as DLB, vascular dementia or normal pressure hydrocephalus. The facial differentiation of DLB and AD patients may be an interesting study in the future. Finally, the facial images included in this study were only front facing images of Japanese participants with neutral expressions. The ability to use images from various angles, of people from varied ethnicities and with a variety of facial expressions would improve the robustness of the AI system.

## CONCLUSIONS

The study showed that deep learning software such as Xception has the ability to differentiate facial images of people with mild dementia from those of people without dementia. This may pave the way for the clinical use of facial images as a biomarker of dementia.

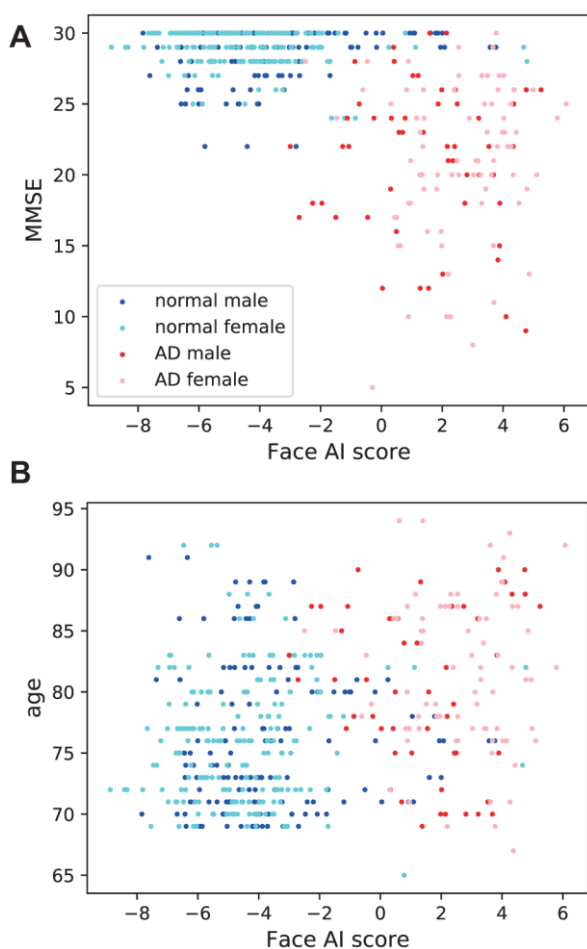
## MATERIALS AND METHODS

### Participants

Dementia patients were recruited mainly from the Department of Geriatric Medicine, The University of Tokyo Hospital. Many of them also participated in a previous perceived age study [5]. The majority of participants without dementia were recruited from the Kashiwa cohort organized by the Institute of Gerontology, The University of Tokyo. Patients from the Department of Geriatric Medicine who were diagnosed as cognitively normal after several tests were included in the non-dementia arm of the study. Participants from the Kashiwa cohort who showed declining MMSE over the past few years were classified into the cognitive impairment arm of the study. The non-AD patients were excluded from the analyses. The rest of the participants were diagnosed to be Alzheimer's disease based on DSM-IV criteria and their Hachinski ischemic scale [10] were no more than 4.

Most of the patients were diagnosed using psychological tests, information from family, laboratory data, brain structural imaging (X-ray computed tomography or nuclear magnetic resonance imaging) and perfusion single photon emission tomography by dementia specialist, except for participants from Kashiwa cohort.

All procedures were approved by the Ethical Review Board at The University of Tokyo Hospital and The



**Figure 3.** Association of Face AI score with (A) MMSE and (B) chronological age. Face AI score correlated closely with (a) MMSE ( $r = -0.599$ ,  $t = -16.40$ ,  $p = 2.47 \times 10^{-48}$ ) and relatively weakly with (b) age ( $r = 0.321$ ,  $t = 7.44$ ,  $p = 4.57 \times 10^{-13}$ ). The Steiger's test found the difference in correlation coefficients to be significant ( $p = 3.25 \times 10^{-35}$ ).

University of Tokyo. The clinical study guidelines of the University of Tokyo, which conform to the Declaration of Helsinki (2013), were strictly adhered to. Healthy volunteers, dementia patients and their families were provided with detailed information about the study, and all provided written informed consent to participate.

### Image preparation

Front-on portrait images were taken of participants wearing a neutral expression. The images were cropped to a square with the face in the middle of the image. Backgrounds were removed so that the AI does not use the background to differentiate cognitive impairment patients from non-dementia patients.

Images of cognitive impairment and healthy participants were divided into 10 groups (group0 ... group9). All images taken of the same participant were included in the same group.

### Selection of models

The network was built using an open-source neural network library, Keras with symbolic tensor manipulation framework, TensorFlow (Google, Mountain View, CA, USA) as a back-end. 5 deep learning models (Xception [11], SENet50 [12], ResNet50 [13], VGG16 [14], simple CNN) with 2 optimizers (SGD, Adam) were tested. Simple CNN was the same network used for the SPECT image study [7]. VGG16, ResNet50, SENet50 were pretrained with VGG-Face [15], and Xception was pretrained with ImageNet [16]. The default settings and parameters for SGD (learning rate = 0.01, momentum=0.9, decay=0.0, nesterov=False) and Adam (learning rate = 0.001,  $\beta_1=0.9$ ,  $\beta_2=0.999$ , =None, decay =0.0, amsgrad=False) were used when being used as optimizers for the simple CNN, but the learning rate was reduced when using the other pretrained models; SGD (learning rate=0.0002, momentum=0.9, decay=0.0, nesterov=False), Adam (learning rate=0.00001,  $\beta_1=0.9$ ,  $\beta_2=0.999$ , =None, decay=0.0, amsgrad=False). Earlier (about 2/3) layers of Xception, SENet50, ResNet50, and VGG16 except batch normalization layers were frozen. All the layers of simple CNN were trainable. Training image data were augmented (rotation range=15, height shift range=0.03, width shift range=0.03, shear range=5, zoom range=0.1, horizontal flip=True, vertical flip=False, brightness range=[0.3, 1.0], channel shift range=5).

5 models with 2 optimizers were tested with group-base 10-fold cross validation. Learning curves were depicted

for 200 epochs. The best model was chosen and the optimum number of epochs determined by considering the accuracy / loss and stability of each model.

### Statistics

The diagnostic and predictive accuracy of the best CNN model was calculated using the group-base 10-fold cross validation. The binary differentiation of cognitive impairment / non-dementia facial images were expressed as "Face AI score". The scores were obtained by applying an inverse sigmoid function to the output predictive value. Face AI score was evaluated using the ROC curve analysis and AUC on the image-base. Correlations between Face AI score and MMSE / chronological age were assessed. Difference by sex was also examined.

All statistical analyses were performed with python and scipy.stat library.

### Analysis of upper and lower half faces

The faces were divided into upper and lower half faces. Upper half faces and lower half faces were separately trained with the same model as the total faces using the same group-base 10-fold cross validation and the optimum number of epochs determined. The performance was analyzed as described above.

### AUTHOR CONTRIBUTIONS

Y.U-K. contributed to conceptualization, dementia patient recruitment, funding acquisition, investigation, and project administration. M.K. contributed to programming deep-learning programs, analyzing data, and writing the initial draft. T.T., B-K.S., and K.I. contributed to participant recruitment in Kashiwa cohort and establishment of the protocol of Kashiwa cohort study. T.K. managed dementia patient recruitment. M.F. and T.I. advised on the construction of the AI models. S.O. and M.A. contributed to supervision and funding acquisition. All the authors discussed the project and have read and approved the final manuscript.

### ACKNOWLEDGMENTS

The authors would like to thank Ms. Yumiko Ishikawa, for evaluating the cognitive scale, taking photographs, and rating; Ms. Mayumi Namba, Ms. Chihiro Tobe, Ms. Hiromi Saito, Ms. Miki Yoshitake, Ms. Yuka Noushige, Ms. Yuiko Taniguchi, and Ms. Ayaka Kashima, for evaluating the cognitive scale; Dr. Masayuki Honda, Dr. Kiwami Kidana, and Dr. Kiyoshi Yamaguchi at the Department of Geriatric Medicine, The University of Tokyo for their help in investigation; Ms. Mariko Maeda

and Mr. Koichi Fujisawa at Sunstar Co., Ltd. for helping taking photographs in the Kashiwa cohort; and Ms. Natalie Okawa for editing the English of this manuscript.

## CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest. The Kashiwa study was also supported by Sunstar Co., Ltd.

## FUNDING

This study was supported in part by Research and Development Grants for Dementia Grant number 16dk0207027h0001, 18dk0207027h0003 and 19dk0207027h0004, provided by Japan Agency for Medical Research and Development (AMED) to M.A. and Grants-in-Aid for Scientific Research (Kakenhi), Grant number 20K07778, provided by Japan Society for the Promotion of Science to Y.U-K, Grant Number 18H05297, to K.I. The Kashiwa study was also supported by Sunstar Co., Ltd., however, Sunstar has no role in this study.

## REFERENCES

1. Christensen K, Thinggaard M, McGue M, Rexbye H, Hjelmberg JV, Aviv A, Gunn D, van der Ouderaa F, Vaupel JW. Perceived age as clinically useful biomarker of ageing: cohort study. *BMJ*. 2009; 339:b5262. <https://doi.org/10.1136/bmj.b5262> PMID:20008378
2. Debrabant B, Soerensen M, Christiansen L, Tan Q, McGue M, Christensen K, Hjelmberg J. DNA methylation age and perceived age in elderly Danish twins. *Mech Ageing Dev*. 2018; 169:40–44. <https://doi.org/10.1016/j.mad.2017.09.004> PMID:28965790
3. Kido M, Kohara K, Miyawaki S, Tabara Y, Igase M, Miki T. Perceived age of facial features is a significant diagnosis criterion for age-related carotid atherosclerosis in Japanese subjects: J-SHIP study. *Geriatr Gerontol Int*. 2012; 12:733–40. <https://doi.org/10.1111/j.1447-0594.2011.00824.x> PMID:22299819
4. Nielsen BR, Linneberg A, Christensen K, Schwarz P. Perceived age is associated with bone status in women aged 25–93 years. *Age (Dordr)*. 2015; 37:106. <https://doi.org/10.1007/s11357-015-9842-5> PMID:26486892
5. Umeda-Kameyama Y, Kameyama M, Kojima T, Ishii M, Kidana K, Yakabe M, Ishii S, Urano T, Ogawa S, Akishita M. Cognitive function has a stronger correlation with perceived age than with chronological age. *Geriatr Gerontol Int*. 2020; 20:779–84. <https://doi.org/10.1111/ggi.13972> PMID:32618098
6. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015; 521:436–44. <https://doi.org/10.1038/nature14539> PMID:26017442
7. Iizuka T, Fukasawa M, Kameyama M. Deep-learning-based imaging-classification identified cingulate island sign in dementia with Lewy bodies. *Sci Rep*. 2019; 9:8944. <https://doi.org/10.1038/s41598-019-45415-5> PMID:31222138
8. Steiger JH. Tests for comparing elements of a correlation matrix. *Psychol Bull* 1980; 87:245. <https://doi.org/10.1037/0033-2909.87.2.245>
9. Fisher RA. On the “probable error” of a coefficient of correlation deduced from a small sample. *Metron*. 1921; 1:1–32.
10. Hachinski VC, Iliff LD, Zilhka E, Du Boulay GH, McAllister VL, Marshall J, Russell RW, Symon L. Cerebral blood flow in dementia. *Arch Neurol*. 1975; 32:632–37. <https://doi.org/10.1001/archneur.1975.00490510088009> PMID:1164215
11. Chollet F. Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017, pp. 1251–1258. <https://doi.org/10.1109/CVPR.2017.195>
12. Hu J, Shen L, Albanie S, Sun G, Wu E. Squeeze-and-excitation networks. *IEEE Trans Pattern Anal Mach Intell*. 2020; 42:2011–23. <https://doi.org/10.1109/TPAMI.2019.2913372> PMID:31034408
13. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016, pp. 770–778. <https://doi.org/10.1109/CVPR.2016.90>
14. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv*. 2014: 1409.1556.
15. Parkhi OM, Vedaldi A, Zisserman A, et al. Deep face recognition. In: British Machine Vision Conference, volume 1. 2015, p. 6. <https://doi.org/10.5244/C.29.41>
16. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg AC, Fei-Fei L. ImageNet Large Scale Visual Recognition Challenge. *Int J Comput Vis*. 2015; 115:211–52. <https://doi.org/10.1007/s11263-015-0816-y>