medRxiv preprint doi: https://doi.org/10.1101/2021.08.17.21262156; this version posted February 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Development and validation of a deep learning algorithm based on fundus photographs for estimating the CAIDE dementia risk score

Rong Hua, Bachelor [1–3]; Jianhao Xiong, PhD[4]; Gail Li, MD, PhD[5]; Yidan Zhu, PhD[1–3]; Zongyuan Ge, PhD[4]; Yanjun Ma, Bachelor [1–3]; Meng Fu, MS[4]; Chenglong Li, Bachelor [1–3]; Bin Wang, MS[4]; Li Dong, MD[6]; Xin Zhao, MS[4]; Zhiqiang Ma, MD[7]; Jili Chen, Bachelor[8]; Chao He, PhD[4]; Zhaohui Wang, MD[7]; Wenbin Wei, PhD[6]; Fei Wang, MD[9]; Xiangyang Gao, MD[9]; Yuzhong Chen, PhD[4]; Qiang Zeng, MD[9]; Wuxiang Xie, PhD[1–3]

R. Hua and J. Xiong contributed equally to this work.

Affiliations:

[1] Peking University Clinical Research Institute, Peking University First Hospital, Beijing, China;

[2] PUCRI Heart and Vascular Health Research Center at Peking University Shougang Hospital, Beijing, China;

[3] Key Laboratory of Molecular Cardiovascular Sciences (Peking University), Ministry of Education, Beijing, China;

[4] Beijing Airdoc Technology Co., Ltd. Beijing, China;

[5] Departments of Psychiatry and Behavioral Sciences and Division of Gerontology and Geriatric Medicine, University of Washington, Seattle, WA, USA.

[6] Beijing Tongren Eye Center, Beijing Tongren Hospital, Beijing, China;

[7] iKang Guobin Healthcare Group Co., Ltd, Beijing, China;

[8] Shibei Hospital Jingan District, Shanghai, China.

[9] Health Management Institute, The Second Medical Center & National Clinical Research Center for Geriatric Diseases, Chinese PLA General Hospital

Corresponding Authors:

Wuxiang Xie, PhD, Peking University Clinical Research Institute, Peking University First Hospital, No. 38 Xueyuan Road, Haidian District, 100191, Beijing, China. Telephone: +86-10-82805564-622; E-mail: xiewuxiang@hsc.pku.edu.cn

Qiang Zeng, MD, Health Management Institute, The Second Medical Center & National Clinical Research Center for Geriatric Diseases, Chinese PLA General Hospital, No. 28 Fuxing Road, Haidian District, 100853, Beijing, China. Telephone: +86-10-68295751; E-mail: zq301@126.com

Total word count of the manuscript: 3492 words. 3 tables, 2 figures, and 1 supplementary file.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.08.17.21262156; this version posted February 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Abstract

Background: The Cardiovascular Risk Factors, Aging, and Incidence of Dementia (CAIDE) dementia risk score is a recognized tool for dementia risk stratification. However, its application is limited due to the requirements for multidimensional information and fasting blood draw. Consequently, effective, and noninvasive tool for screening individuals with high dementia risk in large population-based settings is urgently needed.

Methods: A deep learning algorithm based on fundus photographs for estimating the CAIDE dementia risk score was developed and internally validated by a medical check-up dataset included 271,864 participants, and externally validated by two independent datasets, one included 19,178 medical check-up participants, another included 1,512 community residents. The performance for identifying individuals with high dementia risk (CAIDE score ≥10 points) was evaluated by area under the receiver operating curve (AUC) with 95% confidence interval (CI).

Findings: The algorithm achieved an AUC of 0·944 (95% CI, 0·939–0·950) in the internal validation, 0·877 (95% CI, 0.847–0.907) and 0·781 (95% CI, 0·748–0·814) in the external validations, respectively. Besides, the estimated CAIDE score was significantly associated with both comprehensive cognitive function and specific cognitive domains.

Interpretation: This algorithm trained via fundus photographs could well identify individuals with high dementia risk in a population setting. Therefore, it has potential to be utilized as a noninvasive and more expedient method for dementia risk stratification.

Funding: We were supported by National Natural Science Foundation of China (project no. 81974489), 2019 Irma and Paul Milstein Program for Senior Health Research Project Award, National Key R&D Programme of China (2017YFE0118800).

medRxiv preprint doi: https://doi.org/10.1101/2021.08.17.21262156; this version posted February 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Research in context

Evidence before this study

The retina is an exceptional site where the microcirculation can be handily and noninvasively visualized by fundus photography, thus providing insights into the brain microvasculature. The emerging artificial intelligence technique might be a promising tool to integrate multiple retinal features for identifying individuals with high dementia risk. We searched PubMed up to Feb 24, 2022 with no language restrictions, by the search terms: ("retina" or "fundus") and ("deep learning" or "artificial intelligence" or "AI") and ("dementia" or "Alzheimer's disease" or "CAIDE"), 15 records were yielded. However, we did not find any artificial intelligence algorithm trained by retinal images for estimating or predicting dementia risk.

Added value of this study

To the best of our knowledge, the present study is the first investigation on developing a deep learning algorithm based on fundus photographs for identifying individuals with high dementia risk. The algorithm developed by fundus photographs from 258,305 check-up participants could well identify individuals with high dementia risk, with an AUC of 0·944 in internal validation, 0·877 and 0·781 in two independent external validation datasets, respectively. Besides, the estimated CAIDE dementia risk score exhibited significant association with cognitive function. These findings suggested that the deep learning algorithm based on fundus photographs has potential to identify individuals with high dementia risk in population-based settings. Previous studies have investigated deep learning algorithm based on fundus photographs for predicting cardiovascular diseases, our study added novel evidence regarding dementia in this field, potentially facilitating the eventual application of fundus photography for simultaneous screening of multiple diseases in large population-based settings.

Implications of all the available evidence

This work indicated that a deep learning algorithm trained via fundus photographs could well identify individuals with high dementia risk. Therefore, it has potential application in community-based screening or clinic, and could also be adopted in dementia clinical trials, incorporated as inclusion criteria to efficiently select eligible participants. Future research on escalating the artificial intelligence technology, as well as collecting larger and more detailed datasets, are warranted to further improve and verify the algorithm's performance.

medRxiv preprint doi: https://doi.org/10.1101/2021.08.17.21262156; this version posted February 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Introduction

Worldwide, the number of people have dementia is projected to triply increase to 152 million by 2050, given the dramatic rise in ageing populations, yet there are no curative therapeutics available.[1] Dementia has a long preclinical phase when no symptomatic cognitive impairments, but neurodegenerative progressions are occurring.[2] Early identification of high-risk individuals is essential for preventing dementia, which efficiently targets participants who could benefit most from more intensive examinations and interventions.[3]

The Cardiovascular Risk Factors, Aging, and Incidence of Dementia (CAIDE) dementia risk score was a recognized model to predict 20-year dementia risk, which based on multidimensional risk factors: age, sex, educational level, physical inactivity, systolic blood pressure (SBP), total cholesterol (TC), and body mass index (BMI). It was also highly predictive in external validation of a large multiethnic population,[4-6] and adopted in Finnish Geriatric Intervention Study (FINGER) to select eligible at-risk participants.[7] However, the CAIDE dementia risk score entails measurements by questionnaire inquiry, physical examinations and fasting blood draw, these procedures are time-consuming or invasive for participants, also increase the labor costs of healthcare practitioners and produce biohazardous waste. Consequently, effective, convenient and noninvasive tool to screen individuals with high dementia risk in large population-based settings is warranted.

Vascular disease, especially microvasculature damage in the brain, is recognized as a major contributor to dementia.[1,8] Anatomically and developmentally, the retina shares homology with the brain.[9] The retina is an exceptional site where the microcirculation can be handily and noninvasively visualized by fundus photography, thus providing insights into the brain microvasculature. Large population studies have demonstrated the correlations between various retinal microvascular abnormalities (such as retinopathy, arteriolar narrowing and venular dilation) and increased risk of dementia.[10-12] Moreover, The emerging artificial intelligence technique, especially deep learning, has realized integrating multiple retinal features from fundus photographs, to provide estimation on vascular risk factors, and prediction on cardiovascular diseases.[13-15] However, to our knowledge, this method has not been investigated on predicting dementia.

Herein, we hypothesized that the deep learning algorithm trained via fundus photographs might help to dementia risk stratification. Due to the insufficient time length to occur enough dementia events in our dataset, the present study aimed to train a deep learning algorithm for estimating the CAIDE dementia risk score thus identifying individuals with high dementia risk, and we proposed that the estimated score generated from the algorithm associated with the cognitive function.

Methods

Study design

This was a cross-sectional study. A deep learning algorithm based on fundus photographs for estimating the CAIDE dementia risk score was developed and internally validated by a medical check-up dataset. Additionally, by two independent datasets, one was a medical check-up dataset derived from a different site, another was a community-based cohort dataset, we externally validated the algorithm's discrimination on individuals with high dementia risk. We also further explored the association between the estimated CAIDE dementia risk score and cognitive function based on the community cohort dataset.

Participants and datasets

medRxiv preprint doi: https://doi.org/10.1101/2021.08.17.21262156; this version posted February 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

For the algorithm development, a dataset from 271,864 participants from Tongren Hospital in Beijing, Shibei Hospital in Shanghai, and iKang Healthcare Group who attending medical check-up in 19 province-level administrative regions of China during September 2018 to December 2019, were randomly divided into development (95%) and internal validation (5%) components. This dataset contained retinal fundus images and routine medical information, including age, sex, SBP, TC, and BMI. The use of the dataset for the algorithm training was approved by Tongren Hospital Institutional Review Board, Shibei Hospital Institutional Review Board, and iKang Healthcare Group Institutional Review Board with a waiver of informed consent. The algorithm's performance was further externally validated by two independent datasets. One was the Health Management Institute (HMI) dataset, which included 19,178 medical check-up participants attended the Health Management Institute of Chinese PLA General Hospital during October 2009 to December 2020. The use of the HMI dataset was approved by Chinese PLA General Hospital Institutional Review Board (ethical review approval number: S2019-131-01), all participants provided written informed consent. Another external dataset based on the baseline data from Beijing Research on Ageing and VEssel (BRAVE), a community-based cohort collecting fundus images and health information of middle-aged and older adults in Shijingshan District, Beijing during October to November in 2019.[16] The BRAVE was approved by the Peking University School Institutional Review Board (ethical review approval number: IRB0001052–19060), all participants have given written informed consent.

A variety of digital nonmydriatic fundus cameras were adopted to obtain fundus images, including Canon CR1/CR2 and Crystalvue FundusVue/TonoVue in the development dataset, Canon CR1 in HMI dataset, and Centervue DRS in the BRAVE. All images were captured using 45º fields of view. All datasets calculated the CAIDE dementia risk score based on the function proposed by Kivipelto et al.[4] However, educational level and physical inactivity were not collected in the development dataset. We imputed the risk score of educational level to the algorithm based on the Sixth National Census,[17] according to the average risk score of educational level among the corresponding sex and age group of the individual. Score of physical inactivity was imputed according to BMI status, those overweight or obese participants (defined as BMI $\geq$24 kg/m^2) were regarded as physical inactive, given that the significant association with physical inactive.[18]

Development of the algorithm

The development dataset consists of a training dataset and a tuning dataset. The training dataset was used to update model parameters during the training stage, and the tuning set was used for model selection. The label for training and testing of the network is given as $y_{\text{CAIDE Score}}$ which is the score summation of risk factors according to the CAIDE dementia risk model.[4]

Our CAIDE algorithm was trained and tested using InceptionResNetV2 architecture on the platform Keras v2·2·2 and the Python scikit-learn package 0·22·2. The open source frameworks platform Keras v2·2·2 was available at https://github.com/keras-team/keras. The source code of InceptionResNetV2 was obtained from https://github.com/keras-team/keras-applications/blob/master/keras_applications/inception_resnet_v2.py. The training and testing of the algorithm were performed using a GTX 1080Ti GPU ×2 (CUDA version 9.0, Nvidia Corp., USA) with a batch size of 64 on an operation system Ubuntu v16·04·6. The model was trained for prediction of the CAIDE score as a regression task. We deployed Mean Absolute Error (MAE) as the loss function to minimize during the training stage by Adam optimizer.[19]

The image data was loaded by using OpenCV version 4·2·0. The data augmentations of random

medRxiv preprint doi: https://doi.org/10.1101/2021.08.17.21262156; this version posted February 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

cropping, random rotation (±30°) and random horizontal flipping were implemented by Keras image augmentation package of data generator. In order to improve the robustness of model performance on varying image quality and photography style. An image normalization method, enhanced domain transformation, was used to map any input image pixel values to a given task distribution.[20] To speed up training and validation, multi-processing and 12 workers were utilized by implementing Keras fit generator function.

Validation of the algorithm

The estimated CAIDE dementia risk score of the participants deprived from mean estimated $y_{\text{CAIDE Score}}$ of both eyes, and the actual dementia risk score was calculated according to the CAIDE model. The goodness of fit of the algorithm was assessed by the coefficient of determination (R^2) in the internal validation dataset and the two external datasets. Besides, the algorithm's discrimination on identifying individuals with high dementia risk was evaluated by area under the receiver operating curve (AUC) with 95% confidence interval (CI) by the pROC package version 1·16·2. Consistent with Sindi et al, dementia risk score ≥10 points were recognized as high dementia risk.[6] The maximum Youden index was applied to determine the optimal cut-off point.

Cognitive assessments

We further explored the associations between the estimated CAIDE dementia risk score and cognitive function based on the BRAVE dataset. The primary cognitive measurement in the BRAVE was the Chinese version of Montreal Cognitive Assessment (MoCA) Basic, a sensitive and validated cognitive test battery to comprehensively assess nine cognitive domains.[21] In addition, the BRAVE also supplemented three tests to further assess specific cognitive domains. Specifically, the memory function was measured by immediate and delayed recall of a list of ten unrelated words, and the total score ranged from 0 to 20.[22] The language and executive function was assessed by a verbal fluency test, which requiring participants to speak names of animals as many as possible within 1 minute, and the total number of animal names (excluding repetitive names) was count as the test score.[23] The attention function and executive function were evaluated by the Chinese version of Trails Making Test (TMT),[24] which asking individuals to draw a line through 25 numbers consecutively in ascending order, and as fast as they could. The TMT included two tasks, the TMT-A comprised numbers from 1 to 25, while the TMT-B was different in 25 numbers enclosed in squares from 1 to 12 and circles from 1 to 13. The TMT-A evaluated processing speed and visual attention, and the TMT-B assessed executive function by measuring cognitive alternation ability. In both tests of memory and verbal fluency, the higher score indicated better cognitive performance, while in the TMT, the longer time manifested worse performance.

Statistical analysis

The results were presented using percentage for categorical variables and means ± standard deviations (SD) for continuous variables. We ran multiple linear regression models to examine the associations between the estimated CAIDE dementia risk score and different cognitive assessments. The first model only included the estimated score, while the second model adjusted for multiple covariates, which contained marriage status, drinking status, smoking status, depressive symptoms, *APOE* ε4 status, and chronic diseases status. Specifically, marriage status indicated currently married or not. Participants were divided into non-smokers (including ex-smokers) and current smokers. Alcohol consuming was defined as drinking at least once per week over the past one year. The BRAVE employed the ten-item version of the Center for Epidemiologic Studies Depression Scale (CES-D) to assess depressive

medRxiv preprint doi: https://doi.org/10.1101/2021.08.17.21262156; this version posted February 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

symptoms, with a summed score ranged from 0 to 30. According to the prior study, a score ≥12 was defined as having depressive symptoms in our study.[25] Individuals were divided into *APOE* ε4 carriers (indicated the presence of one or two ε4 alleles) and noncarriers. Diabetes was defined as HbA_{1c} ≥6·5% or fasting blood glucose ≥7·0 mmol/L, or self-reported current use of anti-diabetic therapy. Chronic disease measures also included self-reported physician-diagnosed coronary heart disease, cancer, stroke, and chronic obstructive pulmonary disease. Besides, we also employed analysis of covariance to compare cognitive performance between quartiles of the estimated dementia risk score, with the lowest quartile as the reference. Linear trend was also tested by including risk score quartiles as numerical variables.

To test the robustness of the algorithm, we evaluated the performance of the algorithm using 9 points as the cut-off score of high dementia risk, in consistent with a previous study.[26] We further tested the ability of the algorithm to identify participants eligible for multidomain intervention, since the FINGER trial adopted CAIDE score ≥6 points as one of the inclusion criteria to select eligible at-risk participants among the general population.[7] In addition, we conducted subgroup analyses according to sex, age group (<60 years and ≥60 years), respectively, based on the BRAVE. For algorithm performance in identifying high risk individuals (with CAIDE score ≥10 points), we used Delong test to compare the AUC between subgroups. For the association with cognitive function, we respectively included the interaction terms of estimated dementia risk score with sex, as well as age group in multivariate linear regression models. To investigate the influence of imputation (scores of educational level and physical inactivity) on the algorithm's performance, we additionally developed an algorithm for estimating CAIDE risk score without imputation (which contained scores of age, sex, SBP, TC, and BMI). We combined this algorithm and the actual scores of educational level and physical inactivity in the external validation dataset into an integrated estimated CAIDE dementia risk score, and assessed its performance based on the BRAVE.

All statistical analyses were performed by SAS 9·4 (SAS Institute, Cary, NC), and R language 4·0·0 (R Foundation, Vienna, Austria), with two-tailed alpha value of 0·05 as the statistically significant level.

Role of the Funding Source

The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Results

Study population

The characteristics of individuals in the development dataset, internal validation dataset, and the BRAVE were summarized in Table 1.

Among the 271,864 check-up participants, we randomly divided 95% (258,305 participants, mean aged 42·1 ± 13·4 years, men: 52·7%) into the development group and 5% (13,559 participants, mean aged 41·2 ± 13·3 years, men: 52·5%) into the internal validation group (eFigure 1a). These two groups shared similar baseline characteristics as shown in Table 1. Besides, the characteristics of participants in the training and tuning groups were displayed in eTable 1. A total of 19,178 individuals in the HMI dataset (mean aged 47·8 ± 7·9 years, men: 68.4%) were included in the external validation (eFigure 1b). Among 1,554 individuals taking participant in the baseline survey of BRAVE, 1,512 participants (mean

medRxiv preprint doi: https://doi.org/10.1101/2021.08.17.21262156; this version posted February 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

aged 59·8 ± 7·3 years, men: 37·1%) had fundus photographs and complete information for calculating CAIDE dementia risk score and thus were included in the external validation (eFigure 1c). Among the three datasets, individuals in the BRAVE were older, had a higher proportion of female, with higher SBP. Respectively, 200 (1·5%) individuals in the internal validation dataset, 77 (0·40%) in the HMI dataset, and 159 (10·5%) in the BRAVE were in high dementia risk, with their CAIDE dementia risk score ≥10 points.

Algorithm performance

The R^2 between the estimated and actual CAIDE dementia risk score was 0·80 in the internal validation dataset, 0·54 in the HMI dataset, and 0·32 in the BRAVE (Figure 1). As shown in Figure 2, the algorithm achieved an AUC of 0·944 (95% CI, 0·939–0·950) in the internal validation dataset, 0·877 (95% CI, 0·847–0·907) in the HMI dataset, and 0·781 (95% CI, 0·748–0·814) in the BRAVE for identifying individuals with high dementia risk. The maximum Youden index on the two receiver operating characteristic curves were 0·801 with the sensitivity of 0·959 and specificity of 0·842, corresponded to the optimal cut-off point of 6.793 in the internal validation dataset, 0·624 with the sensitivity of 0·922 and specificity of 0·702, corresponded to the optimal cut-off point of 5.772 in the HMI dataset, and 0·442 with the sensitivity of 0·792 and specificity of 0·650, corresponded to the optimal cut-off point of 8.305 in the BRAVE, respectively.

The estimated score and cognitive function

Linear regression analyses found that the estimated CAIDE dementia risk score (as continuous variable) was significantly associated with the score of MoCA. As shown in Table 2, 1-point increment of estimated CAIDE dementia risk score was significantly associated with −0·565 (95% CI, −0·673 to −0·457) increment of the MoCA score after multivariable adjustment, which manifested worse comprehensive cognitive performance. Similarly, the higher estimated CAIDE dementia risk score was significantly associated with lower score of memory and verbal fluency test, which indicated poorer performance of memory, language and executive function. The higher estimated score was also significantly associated with longer TMT-A and TMT-B time, which represented worse attention and executive function. The analysis of covariance found that after full adjustment, compared with the lowest quartile, the second, third, and highest quartiles were associated with worse comprehensive cognitive function, with lower MoCA score by −0·989 (95% CI, −1·452 to −0·525), −1·685 (95% CI, −2·158 to −1·212), and −2·247 (95% CI, −2·722 to −1·772), respectively (*P* for linear trend < 0·001, Table 3). Similar trends were also observed in performance of memory test, verbal fluency test, TMT-A and TMT-B.

Sensitivity analysis

As shown in eFigure 2, the algorithm still performed well in screening individuals with high dementia risk when the cut-off score changed to 9 points, with an AUC of 0·947 (95% CI, 0·942–0·951) in the internal validation dataset, 0·874 (95% CI, 0·860–0·888) in the HMI dataset, and 0·750 (95% CI, 0·721–0·779) in the BRAVE. As shown in eFigure 3, the algorithm could also identify participants eligible for multidomain intervention, with an AUC of 0·977 (95% CI, 0·975–0·980) in the internal validation dataset, 0·832 (95% CI, 0·825–0·840) in the HMI dataset, and 0·752 (95% CI, 0·725–0·779) in the BRAVE. Besides, eFigure 4 summarized the algorithm performance in subgroups of the BRAVE. The algorithm presented a higher AUC in female (0·808 vs 0·733, *P* = 0·049), as well as in participants <60 years (0·806 vs 0·703, *P* = 0·009). As eFigure 5 presented, we found no interaction effect of sex or age group on the associations between estimated CAIDE dementia risk score and the score of MoCA,

medRxiv preprint doi: https://doi.org/10.1101/2021.08.17.21262156; this version posted February 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

or other specific cognitive functions. In addition, the R^2 between the integrated estimated CAIDE dementia risk score (calculated as the sum of the estimated score derived from the additional algorithm and the actual scores of educational level and physical inactivity in the external validation dataset) and the actual score was 0·60 in the BRAVE, and the integrated estimated CAIDE dementia risk score achieved an AUC of 0·897 (95% CI, 0·873–0·922) in the BRAVE for identifying individuals with high dementia risk (shown in eFigure 6).

Discussion

To the best of our knowledge, the present study is the first investigation on developing a deep learning algorithm based on fundus photographs for identifying individuals with high dementia risk, with an AUC of 0·944 (95% CI, 0·939–0·950) in the internal validation, 0·877 (95% CI, 0·847–0·907) in the HMI dataset, and 0·781 (95% CI, 0·748–0·814) in the BRAVE. Moreover, the estimated CAIDE dementia risk score exhibited significant associations with both comprehensive and specific domains of cognitive function, which further supported the reasonability of the algorithm. Taken together, our study clarified the feasibility of adopting deep learning algorithm based on fundus photographs to screen individuals with high dementia risk in population-based settings.

The rationale of our work based on the concept that, the retina shares similar morphological features and physiological properties with the brain, and hence provide a unique site to detect changes in microvasculature related to the development of dementia.[9] Previous studies have investigated the associations between a spectrum of retinal vascular abnormalities measured via fundus photography and the risk of dementia.[10-12] However, most studies measured retinal signs by semi-automated software, requiring human identification on the basis of prespecified protocols, which might introduce intra- and inter-variability. Besides, recent systematic reviews indicated that combination of multiple retinal vascular parameters, rather than individual marker, might provide higher prognostic value.[27,28] The present study utilized artificial intelligence technique, which might exhibit notable advantages in these issues. Artificial intelligence operates in absence of human assessment, and even performs superiorly to ophthalmologists in capturing subtle retinal changes that would otherwise fail to attract human attention.[29] With faster, easier, more consistent and precise output, the artificial intelligence reduces variability and human cost, thus enhancing the clinical utility of retinal photography.[30] Moreover, artificial intelligence is able to fully extract and integrate multiple retinal features (including information beyond human existed perception or understanding) that are related to dementia risk.

Participants in the BRAVE were much older, and had a larger proportion of female. The significant demographic heterogeneity between the development dataset and this external validation dataset suggested the algorithm's robustness and promising wider utility. One application scenario for the algorithm is screening individuals with high dementia risk in community. Traditional dementia prediction models requiring cognitive tests or multidimensional risk factors increased application difficulties in population-based settings. By contrast, fundus photography is easy to implement and timesaving. According to our practical experience in BRAVE, an investigator with no background on ophthalmology could take fundus photographs within one minute after a few hours of training. Besides, compared with risk factors like blood lipids or glucose, the retinal images have no requirement for fasting status, with less fluctuation and can be obtained noninvasively, thus facilitating the acceptability and convenience of participants. In addition, the algorithm could also be recommended as an add-on to routine screening for diabetic retinopathy, given that patients with diabetes were significantly

medRxiv preprint doi: https://doi.org/10.1101/2021.08.17.21262156; this version posted February 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

associated with higher risk of cognitive decline and dementia.[31] Moreover, our algorithm has potential utility in assessing pre-test dementia probability for further diagnostic tests in outpatient clinics. Last but not the least, this algorithm might also be adopted in dementia clinical trials, incorporated as inclusion criteria to efficiently target eligible participants, or surrogate outcome which could be observed expediently.[7]

Previous studies have investigated deep learning algorithm based on fundus photographs for screening cardiovascular diseases and anaemia,[13,14,32] our study added the novel evidence regarding dementia in this field, potentially facilitating the eventual application of fundus photography for simultaneous screening of multiple diseases in large population-based settings. The foremost strength of the present work was employing convolutional neural network to deal with large dataset of fundus images. The development dataset contained 579,880 fundus images of 258,305 individuals from 19 province-level administrative regions of China, the convolutional neural network exhibited distinct advantages in processing such large dataset, by extracting multiple information from images with a deep architecture, which was similar to image process in human brain.[33] Another strength was incorporating external validation cohorts with varied demographic characteristics and comprehensive cognitive tests, the results externally validated the performance and further supported the scientificalness of the algorithm.

There were, however, also limitations in our study. First, the CAIDE dementia risk score was derived from cross-sectional data, investigations based on incident dementia events in longitudinal settings are warranted to further verify the predictive ability of the algorithm. Second, the R^2 in the BRAVE was relatively small, probably due to the distinct age difference between the development and the BRAVE, given that age is the most important factor for dementia and cognitive function. Another reason could be the absence of educational level and physical inactivity in the development dataset. The sensitivity analysis showed that the integrated estimated CAIDE dementia risk score yielded higher R^2 and AUC. Therefore, future collection of more detailed information in the development dataset could improve the algorithm's performance. Third, the present study only included Chinese participants, which might limit the generalization of our algorithm to other ethnicities.

Conclusions

The present study demonstrated that a deep learning algorithm based on fundus photographs could well identify individuals with high dementia risk, and hold promise for wider application in community-based screening or clinic. As far as we know, this work is the first attempt to utilize deep learning technology and fundus photographs for screening dementia, future advancements in artificial intelligence technology and larger collection of relevant data would further improve and verify the performance of the algorithm.

medRxiv preprint doi: https://doi.org/10.1101/2021.08.17.21262156; this version posted February 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Contributors: WX was responsible for the concept and design. RH, JX, ZG, MF, BW, XZ, CH, YC, LD, ZM, ZW, WW, WF, XG, and WX were responsible for data acquisition, cleaning and interpretation. RH, YZ, YM, CL developed the data analysis plan and preformed the analysis. RH and JX drafted the first manuscript, LG, YC, QZ and WX provided critical revision. WX and QZ obtained the funding. All the authors had full access to all the data and approved the submission of the final manuscript.

Conflict of Interest Disclosure: JX, ZG, MF, BW, XZ, CH, and YC are employees of Beijing Airdoc Technology Co., Ltd. All other authors declare no competing interests.

Data Sharing: Individual participant data will be made available upon reasonable request, directed to the corresponding author (WX and QZ). Data can be shared through a secure online platform for research purposes. We applied the open-source machine-learning framework InceptionResNetV2 to do the experiments. Considering that many aspects of the experimental system (like data generation and model training) largely depend on our internal infrastructure, tooling, and hardware, we are unable to publicly release the code in the present stage. However, the experiments and implementation approaches are provided in the methods section.

Acknowledgments: We thank all participants in the development dataset and external validation datasets.

medRxiv preprint doi: https://doi.org/10.1101/2021.08.17.21262156; this version posted February 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Reference

1. Livingston G, Huntley J, Sommerlad A, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *Lancet* 2020; **396**(10248): 413-46.

2. Sperling RA, Aisen PS, Beckett LA, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* 2011; **7**(3): 280-92.

3. Kivipelto M, Mangialasche F, Ngandu T. Lifestyle interventions to prevent cognitive impairment, dementia and Alzheimer disease. *Nature Reviews Neurology* 2018; **14**(11): 653-66.

4. Kivipelto M, Ngandu T, Laatikainen T, Winblad B, Soininen H, Tuomilehto J. Risk score for the prediction of dementia risk in 20 years among middle aged people: a longitudinal, population-based study. *Lancet Neurol* 2006; **5**(9): 735-41.

5. Exalto LG, Quesenberry CP, Barnes D, Kivipelto M, Biessels GJ, Whitmer RA. Midlife risk score for the prediction of dementia four decades later. *Alzheimers Dement* 2014; **10**(5): 562-70.

6. Sindi S, Calov E, Fokkens J, et al. The CAIDE Dementia Risk Score App: The development of an evidence-based mobile application to predict the risk of dementia. *Alzheimers Dement (Amst)* 2015; **1**(3): 328-33.

7. Ngandu T, Lehtisalo J, Solomon A, et al. A 2 year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): a randomised controlled trial. *Lancet* 2015; **385**(9984): 2255-63.

8. Cheung CYL, Ikram MK, Chen C, Wong TY. Imaging retina to study dementia and stroke. *Prog Retin Eye Res* 2017; **57**: 89-107.

9. Patton N, Aslam T, Macgillivray T, Pattie A, Deary IJ, Dhillon B. Retinal vascular image analysis as a potential screening tool for cerebrovascular disease: a rationale based on homology between cerebral and retinal microvasculatures. *J Anat* 2005; **206**(4): 319-48.

10. Lesage SR, Mosley TH, Wong TY, et al. Retinal microvascular abnormalities and cognitive decline The ARIC 14-year follow-up study. *Neurology* 2009; **73**(11): 862-8.

11. de Jong FJ, Schrijvers EM, Ikram MK, et al. Retinal vascular caliber and risk of dementia: the Rotterdam study. *Neurology* 2011; **76**(9): 816-21.

12. Deal JA, Sharrett AR, Albert M, et al. Retinal signs and risk of incident dementia in the Atherosclerosis Risk in Communities study. *Alzheimers Dement* 2019; **15**(3): 477-86.

13. Poplin R, Varadarajan AV, Blumer K, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nature Biomedical Engineering* 2018; **2**(3): 158-64.

14. Cheung CY, Xu D, Cheng C-Y, et al. A deep-learning system for the assessment of cardiovascular disease risk via the measurement of retinal-vessel calibre. *Nature Biomedical Engineering* 2021; **5**(6): 498-508.

15. Ma YJ, Xiong JH, Zhu YD, et al. Deep learning algorithm using fundus photographs for 10-year risk assessment of ischemic cardiovascular diseases in China. *Science Bulletin* 2022; **67**(1): 17-20.

16. Ma YN, Xie WX, Hou ZH, et al. Association between coronary artery calcification and cognitive function in a Chinese community-based population. *J Geriatr Cardiol* 2021; **18**(7): 514-22.

17. Tubulation on the 2010 population census of the People's Republic of China. 2010. http://www.stats.gov.cn/english/Statisticaldata/CensusData/rkpc2010/indexch.htm (accessed July 22 2021).

18. Tian Y, Jiang C, Wang M, et al. BMI, leisure-time physical activity, and physical fitness in adults

medRxiv preprint doi: https://doi.org/10.1101/2021.08.17.21262156; this version posted February 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

in China: results from a series of national surveys, 2000-14. *Lancet Diabetes Endocrinol*; 2016; **4**: 487–97.

19. DeHoog E, Schwiegerling J. Optimal parameters for retinal illumination and imaging in fundus cameras. *Appl Optics* 2008; **47**(36): 6769-77.

20. Xiong JH, He AW, Fu M, et al. Improve Unseen Domain Generalization via Enhanced Local Color Transformation. International Conference on Medical Image Computing and Computer-Assisted Intervention; 2020.

21. Huang L, Chen KL, Lin BY, et al. Chinese version of Montreal Cognitive Assessment Basic for discrimination among different severities of Alzheimer's disease. *Neuropsych Dis Treat* 2018; **14**: 2133-40.

22. Hua R, Ma YJ, Li CL, Zhong BL, Xie WX. Low levels of low-density lipoprotein cholesterol and cognitive decline. *Science Bulletin* 2021; **16**: 1684-90.

23. Xie WX, Zheng FF, Yan L, Zhong BL. Cognitive Decline Before and After Incident Coronary Events. *Journal of the American College of Cardiology* 2019; **73**(24): 3041-50.

24. Wei MQ, Shi J, Li T, et al. Diagnostic Accuracy of the Chinese Version of the Trail-Making Test for Screening Cognitive Impairment. *J Am Geriatr Soc* 2018; **66**(1): 92-9.

25. Ma Y, Liang L, Zheng F, Shi L, Zhong B, Xie W. Association Between Sleep Duration and Cognitive Decline. *JAMA Netw Open* 2020; **3**(9): e2013573-e.

26. Kaffashian S, Dugravot A, Elbaz A, et al. Predicting cognitive decline: a dementia risk score vs. the Framingham vascular risk scores. *Neurology* 2013; **80**(14): 1300-6.

27. McGrory S, Cameron JR, Pellegrini E, et al. The application of retinal fundus camera imaging in dementia: A systematic review. *Alzheimers Dement (Amst)* 2016; **6**: 91-107.

28. Wagner SK, Fu DJ, Faes L, et al. Insights into Systemic Disease through Retinal Imaging-Based Oculomics. *Transl Vis Sci Technol* 2020; **9**(2): 6-.

29. Son J, Shin JY, Kim HD, Jung KH, Park KH, Park SJ. Development and Validation of Deep Learning Models for Screening Multiple Abnormal Findings in Retinal Fundus Images. *Ophthalmology* 2020; **127**(1): 85-94.

30. Ting DSW, Cheung CYL, Lim G, et al. Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes. *Jama-J Am Med Assoc* 2017; **318**(22): 2211-23.

31. Zheng FF, Yan L, Yang ZC, Zhong BL, Xie WX. HbA(1c), diabetes and cognitive decline: the English Longitudinal Study of Ageing. *Diabetologia* 2018; **61**(4): 839-48.

32. Mitani A, Huang A, Venugopalan S, et al. Detection of anaemia from retinal fundus images via deep learning. *Nature Biomedical Engineering* 2020; **4**(1): 18-27.

33. Anwar SM, Majid M, Qayyum A, Awais M, Alnowami M, Khan MK. Medical Image Analysis using Convolutional Neural Networks: A Review. *J Med Syst* 2018; **42**(11): 226.

medRxiv preprint doi: https://doi.org/10.1101/2021.08.17.21262156; this version posted February 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Figure legends:

Figure 1. Estimation of CAIDE dementia risk score in the internal validation dataset (a), the HMI dataset (b) and the BRAVE (c).

The R^2 (coefficient of determination) between the estimated CAIDE dementia risk score and actual CAIDE dementia risk score is presented.

Abbreviation: HMI = Health Management Institute. BRAVE = Beijing Research on Ageing and Vessel.

Figure 2. Algorithm performance for identifying participants with high dementia risk in the internal validation dataset (a), the HMI dataset (b) and the BRAVE (c).

Individuals with high dementia risk were defined as CAIDE dementia risk score ≥10 points. The points on line indicate the maximum Youden index.

Abbreviation: AUC = area under the receiver operating characteristic curve.

Table 1. Characteristics of individuals in development, internal validation, and two external validation datasets

Characteristics	Development dataset		Internal validation dataset		HMI dataset		BRAVE	
	Men	Women	Men	Women	Men	Women	Men	Women
No. of images	306,795	273,085	16,124	14,246	26,527	12,233	1,122	1,902
No. of participants	136,157	122,148	7,120	6,439	13,119	6,059	561	951
Age (years)	42·3±13·6	41·8±13·8	42·2±13·6	41·9±13·9	48·5±7·8	48·2±8·7	60·7±7·1	59·3±7·3
Systolic blood pressure (mm Hg)	124·1±16·3	116·2±17·8	124·5±16·6	116·1±17·9	125·9±12·8	112·7±19·5	136·1±16·1	130·3±17·4
Total cholesterol (mmol/L)	4·9±0·9	4·9±1·0	4·9±1·0	4·9±1·0	4·8±0·9	4·8±0·9	4·8±0·9	5·0±0·9
Body mass index (kg/m^2)	25·0±3·4	22·8±3·4	25·0±3·5	22·7±3·4	26·2±3·1	23·5±3·2	26·2±3·2	25·8±3·5
Education ≥10 years (%) [a]	-	-	-	-	12,497 (95·2)	5,570 (91·9)	328 (58·5)	576 (60·6)
Physical inactive (%) [b]	-	-	-	-	6,652 (50·7)	3,472 (57·3)	500 (89·1)	739 (77·7)

Data are presented as mean ± SD or n (%).
[a] Educational level was not available in development or internal validation datasets.
[b] Physical inactive was not available in development or internal validation datasets.

medRxiv preprint doi: https://doi.org/10.1101/2021.08.17.21262156; this version posted February 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Table 2. Association between estimated CAIDE dementia risk score and different cognitive assessments: using multiple linear regression models

Cognitive assessments	Model 1[a]		Model 2[b]	
	β **(95% CI)**	***P***	***β*** **(95% CI)**	***P***
MoCA score	−0·589 (−0·692, −0·485)	<0·001	−0·565 (−0·673, −0·457)	<0·001
Memory test score	−0·548 (−0·647, −0·449)	<0·001	−0·511 (−0·603, −0·409)	<0·001
Verbal fluency test score	−0·435 (−0·558, −0·313)	<0·001	−0·415 (−0·542, −0·288)	<0·001
TMT-A time (minutes)	0·099 (0·083, 0·115)	<0·001	0·096 (0·079, 0·112)	<0·001
TMT-B time (minutes)	0·133 (0·115, 0·152)	<0·001	0·132 (0·113, 0·151)	<0·001

[a] Model 1 only included estimated CADIE dementia risk score.
[b] Model 2 adjusted for marriage status, drinking status, smoking status, depressive symptoms, *APOE* ε4 status, diabetes, and self-reported diagnosis of coronary heart disease, stroke, cancer and chronic obstructive pulmonary disease.
Note: The lower score of MoCA, memory test and verbal fluency indicate worse performance. The longer time of TMT-A, TMT-B represent worse performance.

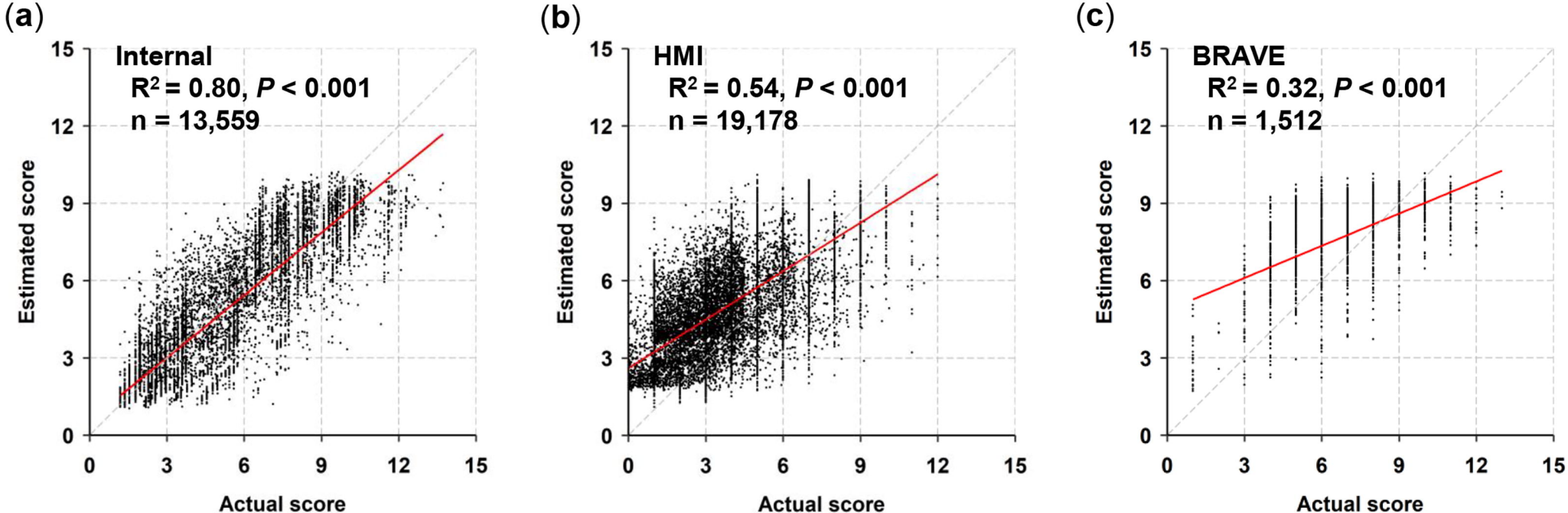
medRxiv preprint doi: https://doi.org/10.1101/2021.08.17.21262156; this version posted February 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

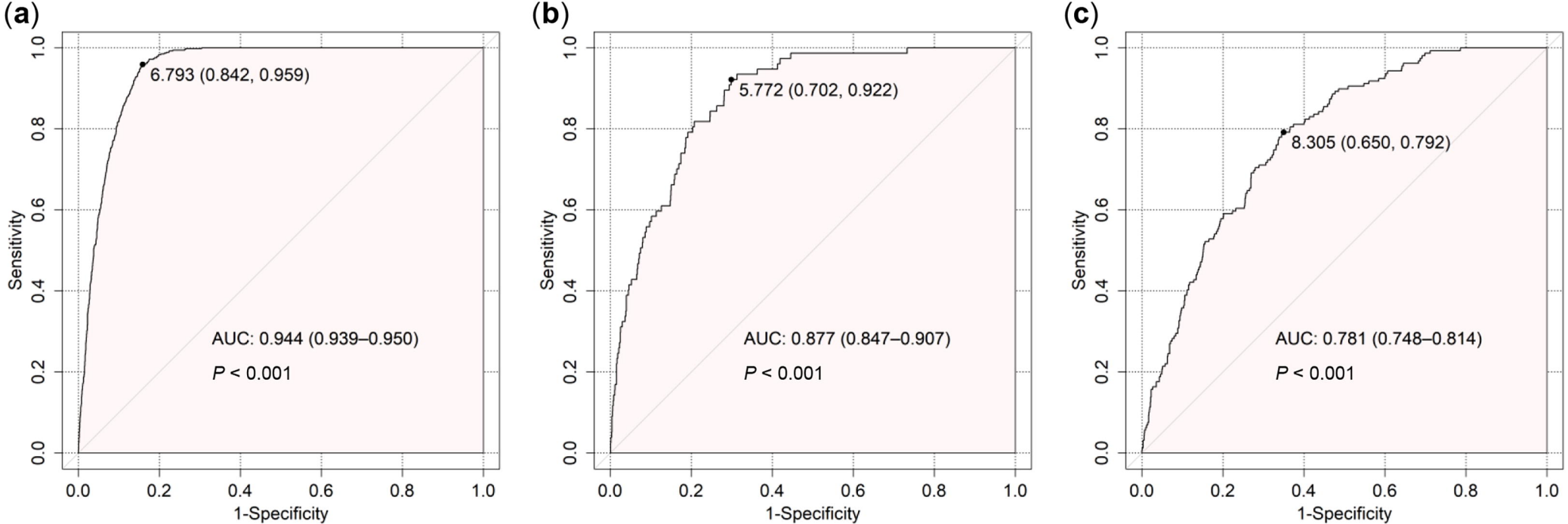
Table 3. Association between quartiles of estimated CAIDE dementia risk score and different cognitive assessments: using analysis of covariance

Estimated CAIDE dementia risk score	MoCA score		Memory test score		Verbal fluency test score		TMT-A time (minutes)		TMT-B time (minutes)	
	β (95% CI)	*P*	*β* (95% CI)	*P*	*β* (95% CI)	*P*	*β* (95% CI)	*P*	*β* (95% CI)	*P*
Model 1[a]										
Quartile 1	0·000 (ref.)	/	0·000 (ref.)	/	0·000 (ref.)	/	0·000 (ref.)	/	0·000 (ref.)	/
Quartile 2	−1·053 (−1·512, −0·594)	<0·001	−1·024 (−1·460, −0·587)	<0·001	−0·561 (−1·101, −0·021)	<0·001	0·127 (0·057, 0·198)	<0·001	0·202 (0·121, 0·284)	<0·001
Quartile 3	−1·778 (−2·237, −1·319)	<0·001	−1·500 (−1·937, −0·063)	<0·001	−1·249 (−1·789, −0·804)	<0·001	0·253 (0·182, 0·323)	<0·001	0·367 (0·285, 0·449)	<0·001
Quartile 4	−2·354 (−2·814, −1·896)	<0·001	−2·220 (−2·656, −1·783)	<0·001	−1·630 (−2·170, −1·089)	<0·001	0·391 (0·320, 0·461)	<0·001	0·510 (0·428, 0·592)	<0·001
Test for linear trend	−0·779 (−0·924, −0·634)	<0·001	−0·713 (−0·852, −0·575)	<0·001	−0·558 (−0·728, −0·387)	<0·001	0·130 (0·107, 0·152)	<0·001	0·169 (0·144, 0·195)	<0·001
Model 2[b]										
Quartile 1	0·000 (ref.)	/	0·000 (ref.)	/	0·000 (ref.)	/	0·000 (ref.)	/	0·000 (ref.)	/
Quartile 2	−0·989 (−1·452, −0·525)	<0·001	−0·917 (−1·356, −0·478)	<0·001	−0·532 (−1·075, 0·011)	0·055	0·122 (0·052, 0·193)	<0·001	0·208 (0·126, 0·290)	<0·001
Quartile 3	−1·685 (−2·158, −1·212)	<0·001	−1·343 (−1·791, −0·895)	<0·001	−1·199 (−1·754, −0·645)	<0·001	0·243 (0·171, 0·316)	<0·001	0·370 (0·287, 0·454)	<0·001
Quartile 4	−2·247 (−2·722, −1·772)	<0·001	−2·065 (−2·515, −1·615)	<0·001	−1·538 (−2·094, −0·981)	<0·001	0·376 (0·303, 0·449)	<0·001	0·502 (0·418, 0·586)	<0·001
Test for linear trend	−0·742 (−0·893, −0·591)	<0·001	−0·662 (−0·805, −0·519)	<0·001	−0·527 (−0·704, −0·350)	<0·001	0·125 (0·102, 0·148)	<0·001	0·167 (0·140, 0·193)	<0·001

[a] Model 1 only included estimated CADIE dementia risk score.
[b] Model 2 adjusted for marriage status, drinking status, smoking status, depressive symptoms, *APOE* ε4 status, diabetes, and self-reported diagnosis of coronary heart disease, stroke, cancer and chronic obstructive pulmonary disease.
Note: The lower score of MoCA, memory test and verbal fluency indicate worse performance. The longer time of TMT-A, TMT-B represent worse performance.

medRxiv preprint doi: https://doi.org/10.1101/2021.08.17.21262156; this version posted February 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.


(a)
Internal
R² = 0.80, P < 0.001
n = 13,559
Estimated score
Actual score
(b)
HMI
R² = 0.54, P < 0.001
n = 19,178
Estimated score
Actual score
(c)
BRAVE
R² = 0.32, P < 0.001
n = 1,512
Estimated score
Actual score


(a)
6.793 (0.842, 0.959)
AUC: 0.944 (0.939–0.950)
P < 0.001
Sensitivity
1-Specificity
(b)
5.772 (0.702, 0.922)
AUC: 0.877 (0.847–0.907)
P < 0.001
Sensitivity
1-Specificity
(c)
8.305 (0.650, 0.792)
AUC: 0.781 (0.748–0.814)
P < 0.001
Sensitivity
1-Specificity