LETTER

ADpred: A non-invasive model for three types of dementia and mild cognitive impairment

Huiyu Xu,[1,2,3,4,7,*] Guoshuang Feng,[5,7] Rong Li,[1,2,3,4] Ping Gao,[6] and Jie Qiao[1,2,3,4]

[1]Center for Reproductive Medicine, Department of Obstetrics and Gynecology, Peking University Third Hospital, Beijing 100191, China
[2]National Clinical Research Center for Obstetrics and Gynecology, Beijing 100191, China
[3]Key Laboratory of Assisted Reproduction (Peking University), Ministry of Education, Beijing 100191, China
[4]Beijing Key Laboratory of Reproductive Endocrinology and Assisted Reproductive Technology, Beijing 100191, China
[5]Big Data Center, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health, Beijing 100045, China
[6]Department of Neurology, Beijing Hospital, Beijing 100006, China
[7]These authors contributed equally
*Correspondence: huiyu56@bjmu.edu.cn

Received: June 22, 2023; Accepted: September 5, 2023; Published Online: September 7, 2023; https://doi.org/10.59717/j.xinn-med.2023.100026

Citation: Xu H., Feng G., Li R., et al., (2023). ADpred: A non-invasive model for three types of dementia and mild cognitive impairment. The Innovation Medicine **1(2)**, 100026.

Dear Editor,

Dementia is a significant neurocognitive disorder that profoundly affects memory, thinking, language, and behavior, resulting in substantial disruptions in daily life. According to the World Alzheimer Report 2010, the aging global population is anticipated to bear a greater economic burden from dementia compared to the combined impact of cancer, heart disease, and stroke.[1] This places a significant strain on healthcare systems and public health resources. While there is currently no cure, available treatments primarily focus on managing symptoms and slowing the progression of dementia. Consequently, the management of risks, early diagnosis, and therapeutic interventions for dementia hold paramount importance.

The diagnosis of dementia and mild cognitive impairment (MCI), a milder form of neurocognitive disorder, relies on a combination of factors, including medical history, examinations, cognitive function assessments, brain imaging, and cerebrospinal fluid (CSF) biomarkers.[2] Among the various types of dementia, Alzheimer's disease (AD) is the most prevalent, followed by vascular dementia (VaD) and dementia with Lewy bodies (DLB).[3] Currently, cognitive assessments are the most convenient and commonly used method for identifying patients. CSF biomarkers, such as amyloid-β (Aβ), total-tau, and phosphorylated-tau, are the most reliable indicators for diagnosing AD.[4,5] However, CSF markers are invasive and typically only useful once clinical cognitive impairment becomes apparent. While some new imaging-based techniques are gaining attention, they are not cost-effective and are unsuitable for early screening of Alzheimer's disease. Ultimately, an ideal biomarker for AD should be non-invasive, user-friendly, cost-effective, and capable of accurately predicting risk and detecting early signs of neurodegenerative processes before clinical cognitive abnormalities become apparent. Currently, there is no existing non-invasive model that can simultaneously screen all three subtypes of dementia and MCI.

Circulating markers offer promising prospects for future dementia screening, primarily due to several advantages they possess. Firstly, they are non-invasive, requiring only a blood sample, thereby eliminating the need for surgeries or expensive imaging procedures. Secondly, the collection and analysis of blood samples are straightforward, cost-effective, and widely applicable, making them feasible and user-friendly. Thirdly, the identification of plasma markers enables early detection of lesions and intervention, potentially slowing down the progression of the disease. However, challenges still remain, such as the reliable selection of markers and the absence of a standardized diagnostic method for dementia using circulating markers. MicroRNAs (miRNAs), which are endogenous non-coding RNAs approximately 22 nucleotides in length, play a crucial role in regulating gene expression at the post-transcriptional level.[6] Due to their small molecular size, plasma miRNAs can traverse the cell membrane and circulate in the blood, making them a valuable tool for non-invasive disease screening of dementia.

GSE12058 provides data on miRNA and other relevant clinical characteristics for predicting dementia. The original paper associated with GSE12058 collected data from individuals with three primary dementia types, as well as individuals with MCI and normal controls. Separate models were developed for each dementia type: AD, VaD, and DLB, utilizing 78 miRNAs, 86 miRNAs, and 110 miRNAs, respectively.[7] Our objective was to utilize this publicly available data to re-establish a model that is straightforward, user-friendly, cost-effective, and widely applicable for all three dementias and MCI. Furthermore, we aimed to transform this model into an online tool that can be accessed and utilized freely.

SUBJECTS ANALYZED

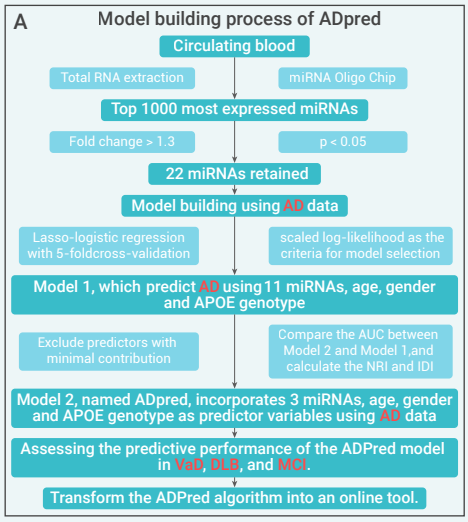
The model-building dataset consisted of 1,309 individuals, including 1,021 AD patients and 288 healthy controls. An additional 91 VaD cases, 169 DLB cases, and 32 MCI cases were included to evaluate the performance of the AD prediction model for other types of dementia and mild neurocognitive disorder. Diagnostic criteria for different types of dementia and definitions of healthy controls are provided in the original data article.[7] Briefly, all 1,601 subjects were over 60 years old, with APOE ε4 genotypes detected and Mini-Mental State Examination (MMSE) assessed. The diagnosis of all patients and healthy controls was based on medical history, physiological examination, diagnostic tests, neurological examination, neuropsychological tests, and brain imaging using magnetic resonance imaging (MRI) or computerized tomography (CT). Neuropsychological tests included MMSE, Alzheimer's Disease Assessment Scale Cognitive Component Japanese version, Logical Memory I and II from the Wechsler Memory Scale–Revised, frontal assessment battery, Raven's colored progressive matrices, and Geriatric Depression Scale. Dopamine transporter imaging and metaiodobenzylguanidine myocardial scintigraphy were used when necessary to diagnose DLB. The AD cases in this study were classified as possible AD, and AD and MCI were diagnosed based on the criteria established by the National Institute on Aging Alzheimer's Association workgroups. VaD and DLB subjects were diagnosed according to the NINDS-AIREN International Workshop and the fourth report of the DLB Consortium, respectively. The MMSE scores of all healthy controls were higher than 23.

The data analyzed in this study was downloaded from the Gene Expression Omnibus (GEO) with the storage number GSE120584, which is accessible online at the following address: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE120584.[7] All the data used in this study were obtained from publicly available resources; therefore, ethical approval was not required.

MODEL CONSTRUCTION

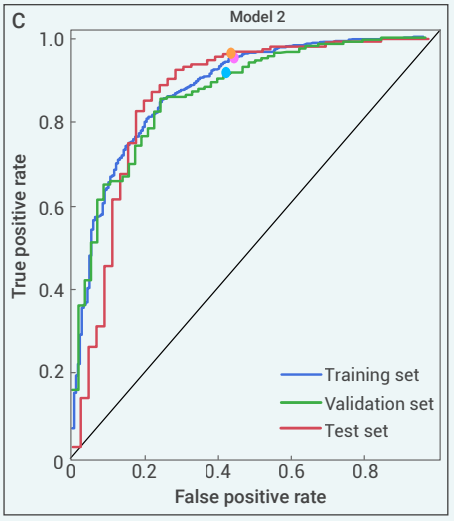
First, the values of normalized miRNA probes were ranked from high to low, and the top 1000 miRNA probes were selected. To identify differentially expressed miRNAs, the Limma program package in R software was utilized. Based on the criteria of $p<0.05$ and fold change > 1.3, a total of 22 miRNAs were ultimately identified.

The flowchart illustrating the model building process is shown in Figure 1A. A cohort consisting of 1,309 individuals diagnosed with Alzheimer's disease (AD) and a normal population was utilized to establish a predictive model for AD. We employed the Least Absolute Shrinkage and Selection Operator (LASSO) logistic regression and performed all statistical analyses using SAS JMPPRO (version 17.0; SAS Institute, Cary, NC, USA). Initially, we randomly allocated 20% of the data as a test set, while the remaining 80% of the data was used for model development. In the model building set, we employed a 5-



B Model 1

Variables	Contribution Total effect) (%)	Contribution Total effect) (%)
Age (years)	53.4	63.0
Gender (male vs female)	3.5	5.6
APOE4ε(+/+ Vs -/-) APOE4ε(+/- Vs -/-)	11.1	17.1
MIMAT0027423	11.0	14.9
MIMAT0019214	2.9	4.5
MIMAT0025845	2.6	4.4
MIMAT0027374	0.8	1.3
MIMAT0024615	0.8	1.5
MIMAT0027425	0.5	0.7
MIMAT0027457	0.3	0.5
MIMAT0027651	0.1	0.2
MIMAT0027382	0.1	0.2
MIMAT0000080	0.1	0.1
MIMAT0019883	1.00E-03	2.00E-03

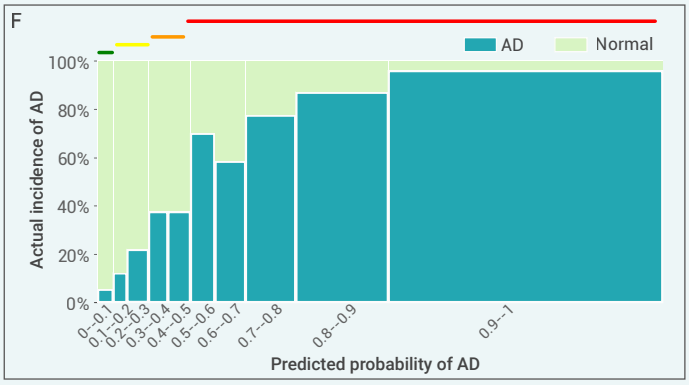


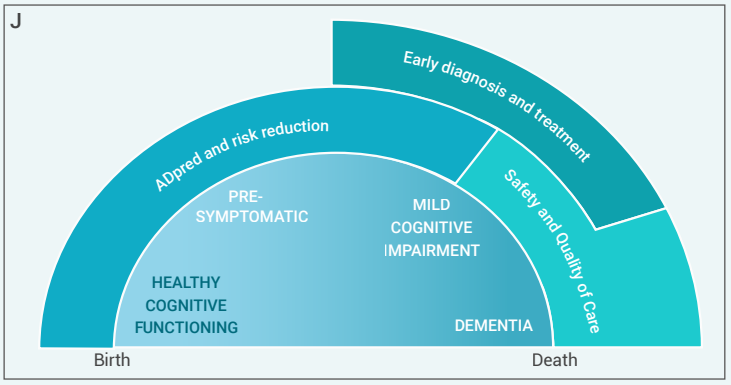
D Model 2

	Training (n=838)	Validation (n=209)	Testing (n=262)
AUC	0.878 (0.856-0.900)	0.875 (0.830-0.920)	0.874 (0.834-0.914)
True Positive Rate	0.953 (0.933-0.966)	0.963 (0.922-0.983)	0.917 (0.871-0.948)
True Negative Rate	0.556 (0.485-0.626)	0.565 (0.422-0.698)	0.579 (0.450-0.698)

E Model 2

Variables	Contribution	
	Main effect (%)	Total effect (%)
Age	52.5	62.1
MIMAT0027423	18.3	24.3
Gender	11.2	17.2
MIMAT0019214	2.6	4.1
MIMAT0025845	2.2	3.5
APOE4ε	1.2	2.3





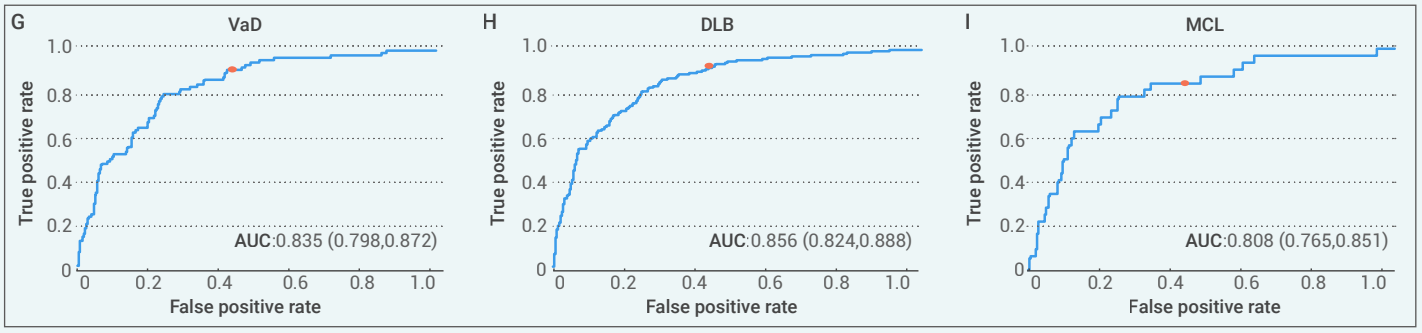


Figure 1. Model building and performance of ADpred (A) Flowchart illustrating the model building process of ADpred. APOE represents Apolipoprotein E. AD denotes Alzheimer's disease, NRI stands for Net Reclassification Improvement, and IDI refers to Integrated Discrimination Improvement. VaD represents vascular dementia, DLB denotes dementia with Lewy bodies, and MCI stands for mild cognitive impairment. (B) Model 1, which predicts AD using 11 miRNAs, age, gender, and APOE genotype. (C) Model 2, created by excluding the 8 miRNAs with minor contributions from Model 1 and utilizing only 3 miRNAs along with other predictors from Model 1. The ROC curves for the training set, validation set, and test set are displayed. (D) The AUCs, true positive rates, and true negative rates of Model 2 in the training set, validation set, and test set are presented. (E) The contribution of each predictor in Model 2. The "Main effect" represents the relative contribution of each variable independently, while the "Total effect" represents the relative contribution of each variable, both individually and in combination with other variables. (F) The correlation between the predicted probability of AD and the actual incidence in the AD dataset, as well as the risk categorization based on this correlation. In the visualization, the low-risk group is represented by the green line, indicating a predicted probability of less than 10%. The yellow line represents predicted probabilities ranging from 10% to less than 30%. The orange line represents predicted probabilities ranging from 30% to less than 50%. Finally, the red line represents predicted probabilities of 50% or higher. (G, H, and I) The ROC curves illustrate the predictive performance of ADpred in VaD, DLB, and MCI, respectively. (J) Potential public health approaches for ADpred-related strategies for dementia.

fold cross-validation method to determine the lambda value in LASSO logistic regression, resulting in the construction of model 1. Model 1 included 11 miRNAs and other predictors, and the contributions of each predictor are displayed in Figure 1B. Subsequently, we eliminated variables with low contributions to model 1 based on variable importance and retained three miRNAs for the reconstruction of model 2. The ROC curves for model 2 in the training

set, validation set, and test set are depicted in Figure 1C, while Figure 1D illustrates the area under the ROC (AUC), true positive rate, and true negative rate of model 2. The variables included and the contribution of each variable in model 2 was demonstrated in Figure 1E. Probe IDs of MIMAT0027423, MIMAT0019214, and MIMAT0025845 were has-miRNA-6761-3p, has-miRNA-3173-5p, and has-miRNA-6716-3p, respectively.

We subsequently compared the AUC of Model 2 with that of Model 1 and calculated the Net Reclassification Improvement (NRI) and Integrated Discrimination Improvement (IDI). The results revealed that there was no statistically significant difference in the AUC between the two models (p=0.639). Additionally, The comparison between the two models also indicated no significant difference in terms of the NRI value and IDI value, which were 0.011 (95% CI: -0.019, 0.041) and 0.006 (95% CI: -0.002, 0.013), respectively.

Figure 1F demonstrates the correlation between the predicted probability of AD using model 2 and the actual incidence of AD. Based on this relationship, individuals were classified into four groups using the predicted probability of AD: Group A represents the lowest-risk category with a predicted AD probability of less than 10%, Group B represents the probability range of 10-30%, Group C represents the probability range of >30-50%, and Group D represents a probability of AD exceeding 50%. Additionally, we have transformed this non-invasive methodology, which utilizes circulating blood, age, and gender, into a freely accessible online tool named ADpred.

Considering that different types of dementia and MCI may share certain common etiological factors, we aimed to evaluate the predictive performances of our ADpred model in forecasting other forms of dementia and MCI, as demonstrated in Figure 1G to 1I. The AUCs for predicting VaD, DLB, and MCI were 0.835 (0.798, 0.872), 0.856 (0.824, 0.888), and 0.808 (0.765, 0.851), respectively. These results suggest that the ADpred model, originally designed to predict AD, also demonstrates good performances for other types of dementia and early cognitive impairment. Furthermore, the findings indicate that the identified risk factors in our model, including advanced age, positive APOE ε4, female gender, and high expression of microRNA MIMAT0027423, are associated not only with an elevated risk of AD dementia but also with other forms of dementia and MCI. Individuals aged over 60 years are encouraged to utilize ADpred for screening AD risk, facilitating early detection of dementia and enabling timely symptom management to prevent disease progression.

MODEL COMPARISON

In our cited research paper (GSE12058), distinct models were developed to predict different subtypes of dementia. The number of miRNAs used in these models varied, with 78 miRNAs for AD, 86 for VaD, and 110 for DLB, respectively.[7] The AUC values for these models in AD, VaD, and DLB were reported as 0.874, 0.867, and 0.870, respectively, in the validation model.[7]

Our constructed ADpred model demonstrates exceptional performance not only in predicting AD but also in predicting VaD dementia, DLB dementia, and MCI. Moreover, it achieves this with the inclusion of only three miRNAs. The AUC values for ADpred in the AD validation data, VaD dementia, DLB dementia, and MCI are 0.874 (95% CI: 0.834, 0.914), 0.835 (95% CI: 0.798, 0.872), 0.856 (95% CI: 0.824, 0.888), and 0.808 (95% CI: 0.765, 0.851), respectively.

Although we utilized AD data for modeling, the AUC of our simplified ADpred model in the validation set aligns with the AUC mentioned in the original text, which is 0.874.[7] However, we did observe a slight decrease in the AUC of ADpred when applied to the VaD and DLB data compared to what was reported in the original paper.[7] Nonetheless, our model underwent significant simplification and achieved satisfactory AUC values across all three dementia types and MCI. Therefore, we firmly believe that ADpred holds even greater practical significance for the early diagnosis of diverse forms of dementia compared to the previous models, which involved a considerably larger number of miRNAs.

LIMITATIONS

First, the population issue, specifically, the high proportion of dementia cases within the data source. In fact, the normal population, even among the elderly, does not exhibit such a high prevalence of dementia. To address this concern, we have proposed a solution: dividing the population into four groups based on the distribution diagram of predicted probability and actual incidence of AD (as depicted in Figure 1F). Secondly, although this model was initially developed for AD dementia and subsequently validated for VaD, DLB, and early cognitive impairment of MCI, it is important to acknowledge that this does not conform to traditional model verifications, additional data collection will be conducted to further validate the model. Lastly, if our model is to be implemented in clinical settings in the future, it will be crucial to develop a more convenient and cost-effective PCR-based miRNA amplification method. This will facilitate the clinical application of our ADpred model.

PERSPECTIVE

AD, VaD, and DLB are the three most common subtypes of dementia. Our established model can predict not only these three subtypes but also MCI, an early stage of cognitive decline. The prediction of MCI is of significant importance, particularly in China where there are 249.49 million individuals aged 60 or older. Among this population, dementia and MCI account for 6.0% and 15.5% respectively.[8] Over a span of 3-5 years, approximately one-third of individuals with MCI remain stable, one-third revert to normal cognition, and one-third progress to dementia.[9] This highlights the presence of underlying mechanisms that either impede or facilitate the progression of MCI, thereby providing an opportunity for intervention and management. Our ADpred tool can assist in the early detection of risk and facilitate early intervention. Figure 1J depicts the potential public health strategies incorporating the use of the ADpred tool for addressing dementia.

RNA therapy, with a focus on untargeted molecular pathways, is an emerging field in biotherapeutics.[10] The miRNAs we have identified in this study hold potential as therapeutic targets for RNA-based drugs, aiming to prevent or slow down the progression of MCI and potentially even offer a cure for dementia.

REFERENCES

1. Chan, K.Y., Wang, W., Wu, J.J., et al. (2013). Epidemiology of Alzheimer's disease and other forms of dementia in China, 1990-2010: A systematic review and analysis. Lancet **381**: 2016–2023. DOI: 10.1016/S0140-6736(13)60221-4.
2. Hugo, J., and Ganguli, M. (2014). Dementia and cognitive impairment: Epidemiology, diagnosis, and treatment. Clin. Geriatr. Med. **30**: 421–442. DOI: 10.1016/j.cger.2014.04.001.
3. Robinson, L., Tang, E., and Taylor, J.P. (2015). Dementia: Timely diagnosis and early intervention. BMJ **350**: h3029. DOI: 10.1136/bmj.h3029.
4. Olsson, B., Lautner, R., Andreasson, U., et al. (2016). CSF and blood biomarkers for the diagnosis of Alzheimer's disease: A systematic review and meta-analysis. Lancet Neurol. **15**: 673–684. DOI: 10.1016/S1474-4422(16)00070-3.
5. Scheltens, P., Blennow, K., Breteler, M.M., et al. (2016). Alzheimer's disease. Lancet **388**: 505-517. DOI: 10.1016/S0140-6736(15)01124-1.
6. Catalanotto, C., Cogoni, C., and Zardo, G. (2016). MicroRNA in control of gene expression: An overview of nuclear functions. Int. J. Mol. Sci. **17**: 10. DOI: 10.3390/ijms17101712.
7. Shigemizu, D., Akiyama, S., Asanomi, Y., et al. (2019). Risk prediction models for dementia constructed by supervised principal component analysis using miRNA expression data. Commun. Biol. **2**: 77. DOI: 10.1038/s42003-019-0324-7.
8. Jia, L., Du, Y., Chu, L., et al. (2020). Prevalence, risk factors, and management of dementia and mild cognitive impairment in adults aged 60 years or older in China: A cross-sectional study. Lancet Public Health **5**: e661–e671. DOI: 10.1016/S2468-2667(20)30185-7.
9. Manly, J.J., Tang, M.X., Schupf, N., et al. (2008). Frequency and course of mild cognitive impairment in a multiethnic community. Ann. Neurol. **63**: 494–506. DOI: 10.1002/ana.21326.
10. Crooke, S.T., Witztum, J.L., Bennett, C.F., et al. (2019). RNA-targeted therapeutics. Cell Metab. **29**: 501. DOI: 10.1016/j.cmet.2019.01.001.

DECLARATION OF INTERESTS

Jie Qiao, Rong Li and Guoshuang Feng are the Editorial Board member of The Innovation Medicine. They were blinded from reviewing or making final decisions on the manuscript. The article was subject to the journal's standard procedures, with peer review handled independently of this Editorial Board member and their research groups. The other authors declare no competing interests.

DATA AND CODE AVAILABILITY

The data analyzed in this study was downloaded from the GeneExpres-sion Omnibus (GEO) with the storage number GSE120584, which isaccessi-ble online at the following address: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE120584.