A Strategy for Identifying the Risk of Developing Alzheimer's Disease From Healthy Controls: Distance to Novelty Detection Boundary

Jiangbing Mao
Fujian Normal University https://orcid.org/0000-0003-0932-0847
Qinyong Ye
Fujian Medical University Union Hospital
Hongqing Yang
Fujian Normal University Cangshan Campus: Fujian Normal University
Magda Bucholc
Ulster University
Shuo Liu
Ulster University
Wenzhao Gao
Ulster University
Yingqing Wang
Fujian Medical University Union Hospital
Jiawei Xin
Fujian Medical University Union Hospital
Xuemei Ding (x.ding@ulster.ac.uk)
Ulster University https://orcid.org/0000-0002-8644-4256

Research

Keywords: Alzheimer's disease, Mild cognitive impairment, Novelty detection, Machine learning, Artificial intelligence, Predictive model

Posted Date: December 17th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-129454/v1

A strategy for identifying the risk of developing Alzheimer's disease from healthy controls: distance to novelty detection boundary

Jiangbing Mao[1, #] , Qinyong Ye[2, #], Hongqin Yang[1,*], Magda Bucholc[3], Shuo Liu[3], Wenzhao Gao[4], Yingqing Wang[2], Jiawei Xin[2], Xuemei Ding[3,*] and the Australian Imaging Biomarkers and Lifestyle flagship study of ageing[$]

[1] Fujian Provincial Key Laboratory for Photonics Technology, Key Laboratory of OptoElectronic Science and Technology for Medicine of Ministry of Education, Fujian Normal University, Fuzhou 350007, China.

[2] Department of Neurology, Fujian Medical University Union Hospital, Fuzhou 350001, China.

[3] Intelligent Systems Research Centre, School of Computing, Engineering and Intelligent Systems, Ulster University, Derry~Londonderry, UK, BT48 7JL.

[4] School of Computing, Engineering and Intelligent Systems, Ulster University, Derry~Londonderry, UK, BT48 7JL.

[#] These authors equally contribute to this work.

* Corresponding authors: Xuemei Ding (x.ding@ulster.ac.uk), Hongqin Yang (hqyang@fjnu.edu.cn).

[$]Data used in the preparation of this article was obtained from the Australian Imaging Biomarkers and Lifestyle flagship study of ageing (AIBL) funded by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) which was made available at the ADNI database (www.loni.usc.edu/ADNI). The AIBL researchers contributed data but did not participate in analysis or writing of this report. AIBL researchers are listed at www.aibl.csiro.au.

Abstract

Background:

Machine learning (ML) techniques are expected to tackle the problem of the high prevalence of Alzheimer's disease (AD) we are facing worldwide. However, few studies of novelty detection (ND), a typical ML technique for safety-critical systems especially in healthcare, were engaged for identifying the risk of developing cognitive impairment from healthy controls (HC) population.

Materials and Methods:

Two independent datasets were used for this study, including the Australian Imaging Biomarkers and Lifestyle Study of Ageing (AIBL) and the Fujian Medical University Union Hospital (FMUUH), China datasets. Multiple feature selection methods were applied to identify the most relevant features for predicting the severity of AD. Four easily interpretable ND algorithms, including k nearest neighbor, Mixture of Gaussian (MoG), KMEANS, and support vector data description were used to construct predictive models. The models were visualized by drawing their decision boundaries tightly surrounding the HC data. A distance to boundary (DtB) strategy was proposed to differentiate individuals with mild cognitive impairment (MCI) and AD from HC.

Results:

The best overall MCI&AD detection performance in both AIBL and FMUUH was obtained on the cognitive and functional assessments (CFA) modality only using MoG-based ND with AUC of 0.8757 and 0.9443, respectively. The highest sensitivity of MCI was presented by using a combination of CFA and brain imaging modality. The DTB value reflects the risk of developing cognitive impairment for HC and the dementia severity of MCI/AD.

Conclusions:

Our findings suggest that applying some non-invasive and cost-effective features can significantly detect cognitive decline in an early stage. The visualized decision boundary and the proposed DtB strategy illustrated the severity of cognitive decline of potential MCI&AD patients in an early stage. The results would help inform future guidelines for developing a clinical decision-making support system aiming at an early diagnosis and prognosis of MCI&AD.

Keywords: Alzheimer's disease, Mild cognitive impairment, Novelty detection, Machine learning, Artificial intelligence, Predictive model

Background

With the ageing population, we are challenged by a growing impact of neurodegenerative diseases, such as Alzheimer's disease (AD), the most common type of dementia [1]. Especially, a steep worsening of neuropsychiatric symptoms of dementia has been reported during the COVID-19 pandemic [2]. Neurodegeneration incrementally diminishes the quality of a patient's life and leads to heavy economic burden in healthcare. In 2019 the Alzheimer's Disease International (ADI) reported that over 50 million people worldwide suffered from dementia. The total estimated worldwide cost of dementia in 2018 was US$ 1 trillion. This figure is projected to double by 2030 [3]. The total number of people with dementia in China was 9.5 million in 2015 and will increase to 14.1 million by 2020 and 23.3 million by 2030 [4]. The total costs of dementia in China have been predicted to reach US$ 69 billion in 2020 and US$ 114.2 billion in 2030 [5].

There is an urgent need to develop AI-enabled clinical diagnosis support systems (AI-CDSS) to accelerate AD diagnosis and prognosis and to improve the quality of healthcare. The AI-CDSSs can outperform the traditional CDSS, improve the clinical decision-making process, reduce medical errors, and decrease costs [6]. The development of AI raised to a governmental strategy level in many countries, including

the US, China, and the UK. For example, the UK Health Secretary recently announced a £250 million investment in establishing a new national AI lab that will use the power of AI to improve the health and lives of patients [7]. AI is widely seen as having the potential to improve efficiency across many sectors, and healthcare is one of the major and the most critical domains to be revolutionized by AI [8].

Novelty detection (ND) aims to identify behaviour that are not consistent with normal expectations, representing a machine learning method in AI that could be incorporated into the AI-CDSS as it is more accurate, efficient and applicable particularly in monitoring for safety-critical systems [9, 10]. ND has gained much attention in application domains including anomalous behaviour detection for elderly care [11][12][13], prediction of new disease-causing genes [14], and cancer image classification [15].

Although ND has been widely applied in medical and healthcare fields, to the best of our knowledge, there is no study evaluating the risk of developing AD using ND techniques. Research in this area has great potential for the prevention and treatment of dementia. Unlike other methods in AI that need balanced data across each given class when building models, ND techniques are applicable when only one (i.e. normal) class data, HC data in this scenario, are available. The produced ND model can be used to detect whether or not healthy elderly adults are at the risk of developing AD at an

early stage, which will in turn be referred by clinicians for follow-up treatment.

This study aims to construct an interpretable novelty detector, which constructs a tightly closed decision boundary to differentiate between healthy controls and patients with dementia. We also propose a novel distance to boundary (DtB) strategy for evaluating the severity of the risk of developing AD of a patient based on the distance of a data point to the decision boundary.

Data and Methods

Data description

Two independent datasets were used in this study, namely, the Australian Imaging Biomarkers and Lifestyle Study of Ageing (AIBL) and the Fujian Medical University Union Hospital (FMUUH) datasets.

The AIBL data

Patient records from the AIBL database (https://www.aibl.csiro.au/) were used in this study to build the ND models. The data was collected by the AIBL study group. AIBL study methodology has been reported previously [16].

The primary goal of the AIBL study is to discover which cognitive characteristics,

biomarkers, and health and lifestyle factors determine the subsequent development of symptomatic AD. The AIBL clinical data were extracted mainly from seven categories: cerebrospinal fluid biomarkers (CSF), cognitive and functional assessments (CFA), magnetic resonance imaging (MRI), Positron emission tomography (PET), blood test (BLO), demographic (DEM), and medical history (MH). In our study, the CFA adopted the Mini-Mental State Examination (MMSE) [17] and logical memory immediate/delayed recall assessments (LMIR/LMDR). The brain imaging data used in the study consisted of coarse-grained structural MRI and PET with [^{11}C]-Pittsburgh compound B (PIB). They are the total volume for grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF), and the total number of active pixels (PIB.PET) from brain imaging. Therefore, we used 32 features as potential predictors of cognitive decline associated with AD. They are 3 CFA, 12 BLO, 4 neuroimaging, 2 sociodemographic, 10 medical history (MH), and ApoE genotype features (Table 1).

Table 1 All features in AIBL data.

Category	Features
Cognitive and functional assessments (CFA)	MMSE, LMIR, LMDR.
Blood analyses (BLO)	Thyroid stimulating hormone (AXT117), vitamin B12(BAT126), red blood cell(HMT3), white blood cell (HMT7), platelets (HMT13), haemoglobin (HMT40), mean corpuscular haemoglobin (HMT100), and mean corpuscular haemoglobin concentration (HMT102);

	Urea nitrogen (RCT6), serum glucose (RCT11), cholesterol high performance, RCT20), creatinine (rate blanked, RCT392).
Brain imaging (IMG)	PIB-PET, MRI (GM,WM,CSF)
Demographics (DEM)	Gender, age.
Medical history (MH)	Psychiatric (MHPSYCH), neurologic other than AD (MH2NEURL), head, eyes, ears, nose and throat disease (MH4CARD), hepatic (MH6HEPAT), musculoskeletal (MH8MUSCL), endocrine-metabolic(MH9ENDO), gastrointestinal (MH10GAST), renal-genitourinary (MH12RENA), smoking (MH16SMOK) and malignancy (MH17MALI) histories.
ApoE genotype	Two alleles apolipoprotein genotype (ApoE).

The Clinical Dementia Rating (CDR) scores were used to characterize the severity of AD and acted as a response measure in prediction models. Accordingly, participants were categorized into five groups based on the CDR scale levels: the healthy controls (CDR = 0), mild cognitively impaired (MCI) (CDR = 0.5), the mild (CDR = 1), moderate (CDR = 2) and severe (CDR = 3) AD patients. CDR has been previously used as an objective measure of the severity of AD and been shown to be highly correlated with the clinical diagnosis [18]. The subjects from mild, moderate, and severe AD categories were combined into one AD category.-Note that only health controls (HC) data were used to train the ND models. The non-healthy data were considered as abnormal data and was used to optimize the model parameters and test the predictive performance of the model.

There were a total of 861 participants at baseline (BL), with 262, 222, and 142 participants had a record of a follow-up visit at 18 (M18), 36 (M36), and 54 (M54) month, respectively. In our study, the data at different times were combined as one dataset to train the ND model. While dataset with complete features (i.e. CFA, MRI, PIB.PET, BLO, DEM, MH and ApoE; 641 subjects in total) was used to perform comparative experiments with different combinations of modalities; the dataset without neuroimaging features (including CFA, BLO, DEM, MH and ApoE; 1487 subjects in total) was used to implement the proposed DtB strategy.

The FMUUH data

In addition to the AIBL data, this study also considered 330 clinical data records (148 HC and 182 AD) obtained from FMUUH. Six features were considered, namely, three cognitive assessments including MMSE, Alzheimer's Disease Cooperative Study - Activities of Daily Living (ADCS-ADL) [19], and the Neuropsychiatric Inventory (NPI)[20], and three demographics (Age, Education Level, and Gender). Detailed analysis of AIBL and FMUUH demographic characteristics can be accessed in Additional file 1: Table S1 and Fig.S1.

Feature selection

Feature selection techniques were applied to minimize the computational costs, decrease analytical complexity, and identify significant features associated with the

severity of AD. First, min-max normalization was conducted to assimilate clinical measurements of diverse scales into the range of 0~1. To avoid bias associated with the choice of one specific feature selection technique for establishing feature ranking, we have used three different univariate feature selection approaches to select the optimal number of features, namely, based on information gain ratio (IGR) [21], Pearson's correlation [22], and Chi-square [23]. The Cross-Entropy Monte Carlo rank aggregation algorithm [24] was then utilized to aggregate the ranking results obtained by the above three filters and finally get the top-10 significant features.

Novelty detection modelling approaches

Novelty detection is also referred as one-class classification [25], anomaly detection [26] or data description [27]. The ND modelling process is described in Fig. 1.

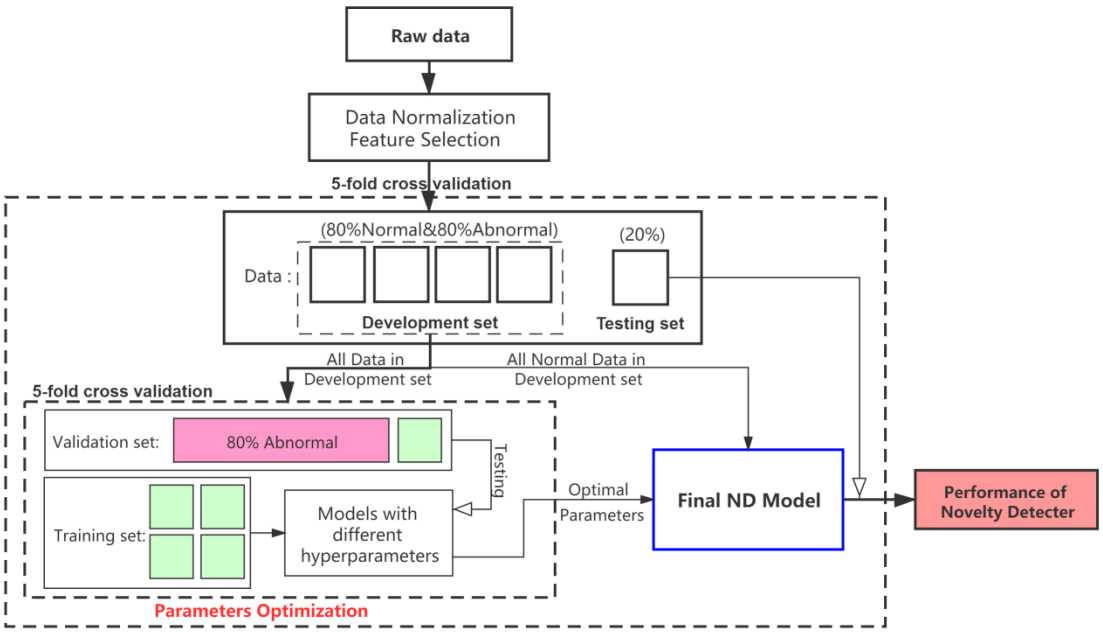

Fig.1 The overall framework of the ND modelling process.

The preprocessed dataset, with respect to each category, was split into five folds: four folds for model development (including 80% normal and 80% abnormal data) and one fold for model testing (20% normal + 20% abnormal). To avoid any bias introduced by randomly partitioning and to get better repeatability, each one of the five folds was selected as the testing set and the remaining four folds were used as a development set. (See the outer 5-fold CV loop in Fig. 1). Next, the normal data in each development set were further split into five folds for model training and validation. The training set only included four-folds normal data which were used to construct an ND model, while the remaining one fold normal data were combined with the 80% abnormal data as a validation set to validate the constructed model. This process was iterated five times (the inner 5-fold CV shown in Fig. 1) to tune the hyperparameters of the model to get an unbiased evaluation of the model fitting on the training set. Finally, the tuned/optimized hyperparameters and the entire normal data in the development set were applied to construct an optimal ND model whose performance would then be assessed using a testing set unseen during the model's training. The experimental results are averaged over five outer folds.

According to [28] and for interpretability, four representative ND methods, namely k nearest neighbor (KNN), Mixture of Gaussian (MoG), KMEANS and support vector data description (SVDD), were employed for ND modelling in this study. The selection of

these methods is due to their comprehensive interpretability, popular applicability in various domains [29], outstanding historical contributions to ND methods development [30], and the potential expandability for further research [31]. Their computational principles are easily understandable and briefly introduced as follows.

KNN

The KNN is a representative distance-based ND method assuming that all normal data points are close to each other, and anomalies are far from the normal points [32]. The KNN method first calculates the distance between the data point x and its k nearest neighbors (denoted as $NN_k(x)$) and then calculates the distance from these nearest neighbor $NN_k(x)$ to their k nearest neighbors $NN_k(NN_k(x))$. Finally, it discriminates whether a data point x is normal or abnormal by comparing these two distances. The acceptance function for a test data point can be defined as [32]:

$$f_{NN_k}(x) = I\left(\frac{\|x - NN_k(x)\|}{\|NN_k(x) - NN_k(NN_k(x))\|} \leq 1\right)$$

where $I(\bullet)$ is an indicator function. If $\bullet$ is true, then $I(\bullet)$ =1 indicates x normal; otherwise $I(\bullet)$=0 indicates x abnormal.

MoG

The MoG is a commonly used density-based ND method by calculating a linear combination of N components of normal distribution on the given data. The

probability density of data x can be estimated with [33]:

$$P_{GM}(x)=\frac{1}{N}\sum_{j}^{N}\left\{a_j\frac{1}{(2\pi)^{d/2}\left|\sum_j\right|^{1/2}}\exp\left\{-\frac{1}{2}(x-\mu_j)^T\sum_j^{-1}(x-\mu_j)\right\}\right\}$$

where a_j is the mixture coefficients, μ_j is the mean of the j[th] Gaussian component, and $\sum_j$ is the covariance matrix. Data lying in a high density area are accepted as normal; otherwise are detected as abnormal.

KMEANS

KMEANS, a representative clustering-based ND method, is one of the most popular techniques due to its simplicity of implementation [33]. This method clusters normal data using a small number (i.e. k) of prototypes. The centroids of k clustered prototypes are optimized by the following minimised square error:

$$error_{k-means}=\sum_i\left(\min_k\|x_i-\mu_k\|^2\right)$$

where μ_k is the centroid associated with the k^{th} cluster. Any data excluded by all clusters would be detected abnormal.

SVDD

The SVDD represents a support vector machine-based ND method. It employs a hypersphere to define a closed decision boundary around normal data. The SVDD method can be simplified as [27]:

$$f_{SVDD}(x_i, a, R) = I\left(\|x_i - a\|^2 \leq R^2 + \xi_i\right)$$

where x_i is a data point from the training set, a is the centre of the hypersphere, and R is the distance from a to the support vectors on the decision boundary. In this study, the radial basis function[35] was used as the kernel of SVDD to map the target data onto a boundary.

Performance metrics

To measure the ND performance, we employed three performance metrics: sensitivity, specificity and the area under the receiver operating characteristic (ROC) curve (AUC) [36]. In the context of ND in the medical domain, positive (MCI/AD) and negative (HC) represents abnormal and normal, respectively.

Sensitivity reflects the rate of the abnormal data correctly detected. Therefore, we used sensitivity to assess the power of a ND model detecting different severity of getting AD of individuals. The higher the sensitivity, the less likely an AD patient fails to be pinpointed in diagnosis and prognosis. While, specificity represents the rate of the normal data correctly detected. A higher specificity indicates that the novelty detector is less likely to misdiagnose healthy controls. AUC is an integrated quantitative presentation of the ROC curve plotted the sensitivity against the 1-specificity at various thresholds. It can be used to measure the overall performance of a novelty detector thoroughly.

Results and Analysis

Feature selection results

To determine if cost-effective and non-invasive AD markers have high discriminative power when they are used for detecting potential AD patients, all features in the AIBL data were grouped into four modalities: 1) CFA including LMDR, LMIR, and MMSE; 2) brain imaging features (IMG); 3) medical history and demographics (MH&DEM); and 4) a two alleles apolipoprotein genotype and blood test (BLO&ApoE).

Fig. 2A shows that the CFA has the highest correlation with CDR, followed by IMG (excluding WM) and ApoE. In contrast, MH&DEM and BLO modalities are less correlated to CDR. The scores of these features' importance are close to zero by applying Chi-square and IGR filters. Different aggregation plans are used to test the ND model performance with/without expensively obtained, invasive test and complex analysis required modalities (e.g. IMG, BLO, and ApoE, respectively). First, we aggregate the feature ranking of all modalities (Fig. 2B). Then, the costly obtained, invasive and complex analysis required features (IMG, BLO, and ApoE) are removed (Fig. 2C&D, respectively). Finally, the top-10 important features were selected after aggregation for the convenience of analysis. Performance of the ND model trained by all the other possible aggregation schemes can be found in Additional file 1: Table S2.

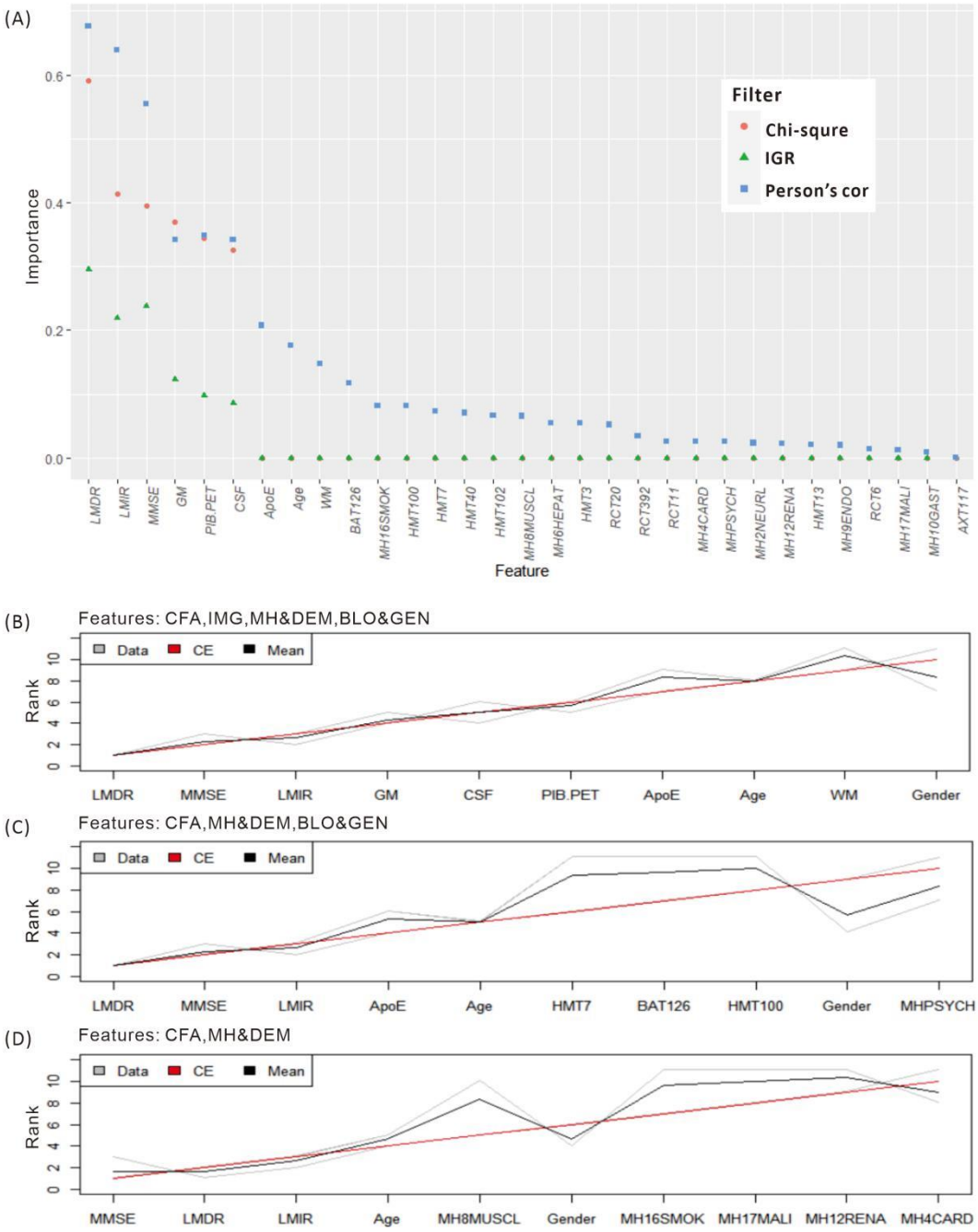

Fig.2 Ranking features importance and aggregating the ranking results with different combinations of modalities. (A). The three filters applied are based on Chi-square value (orange dot), information gain ratio (IGR, green triangle), and Pearson's correlation (blue square). The gray and black lines represent three ranking results and their averages, while the red line reflects the final aggregation result. Some Chi-square values are overlapped by IGR as they all close to zero.

Model performance results

A thorough comparison between the ND models constructed on both datasets was conducted to verify the applicability and robustness of the applied methods and the proposed DtB strategy.

AIBL data

Table 2 compares the AUC, specificity, and sensitivity performance of the ND models produced by KNN, MoG, KMEANS, and SVDD algorithms applied to different modality combinations. It turns out feature selection significantly improved the performance of MoG, KMEANS, and SVDD. Furthermore, the CFA data can improve the performance of all the models. The models performance obtained on all different modality combinations with and without feature selection is included in the supplementary Tables S2 and S3.

Table 2 The performance comparison of the ND models constructed by KNN, MoG, KMEANS, and SVDD using different modality combinations of the AIBL data.

	False rejection rate: 0~1	False rejection rate = 0.1	
ND algorithms	AUC (95%CI)	Specificity	Sensitivity

Modality combinations			HC	MCI	AD
CFA,IMG,MH&DEM,BLO&ApoE (without feature selection)	KNN	0.8479 (0.7897-0.9061)	90.79%	30.25%	38.59%
	MoG	0.7890 (0.6610-0.9169)	82.12%	43.24%	88.47%
	KMEANS	0.7551 (0.6924-0.8178)	67.52%	61.78%	83.57%
	SVDD	0.6856 (0.5604-0.8107)	77.65%	59.26%	73.68%
CFA,IMG,MH&DEM,BLO&ApoE	KNN	0.8690 (0.8240-0.9140)	89.73%	62.64%	92.82%
	MoG	0.8552 (0.8185-0.8919)	85.89%	64.55%	92.82%
	KMEANS	0.8676 (0.8179-0.9173)	87.43%	58.02%	97.39%
	SVDD	0.8462 (0.7636-0.9288)	88.28%	51.73%	89.54%
CFA,MH&DEM,BLO&ApoE	KNN	0.8000 (0.7127-0.8873)	89.59%	54.84%	91.89%
	MoG	0.8266 (0.7624-0.8908)	86.04%	59.14%	91.89%
	KMEANS	0.8257 (0.7329-0.9184)	90.59%	52.08%	86.69%
	SVDD	0.8304 (0.6963-0.9644)	86.50%	56.28%	93.80%
CFA,MH&DEM	KNN	0.7858 (0.6710-0.9005)	89.04%	53.68%	86.84%
	MoG	0.8048 (0.6880-0.9214)	89.02%	47.53%	75.43%
	KMEANS	0.7452 (0.6686-0.8217)	84.08%	52.55%	66.72%
	SVDD	0.7783 (0.6646-0.8919)	86.21%	63.93%	83.99%
CFA,IMG	KNN	0.8464 (0.7380-0.9548)	91.40%	68.61%	92.27%
	MoG	0.8558 (0.7717-0.9400)	90.69%	69.86%	95.85%
	KMEANS	0.8564 (0.7492-0.9636)	90.71%	67.60%	96.87%
	SVDD	0.8442 (0.7190-0.9694)	88.99%	68.57%	92.72%

CFA	KNN	0.8521 (0.7250-0.9792)	89.40%	59.87%	96.79%
	MoG	**0.8757 (0.7982-0.9532)**	**89.63%**	**67.33%**	**96.79%**
	KMEANS	0.8527 (0.7405-0.9650)	89.11%	56.92%	95.23%
	SVDD	0.8267 (0.7013-0.9521)	84.94%	60.83%	98.43%

False rejection rate (i.e., a tolerant error) identifies the fraction of the HC diagnosed as non-healthy. CI: confidence interval.

Interestingly, models built on CFA only performed better than most other modality combinations in terms of the AUC value. Based on the 5-fold CV assessment results, the MoG produced the highest AUC of 0.8757 (95%CI: 0.7982-0.9532) (Table 2, bold), and the KMEANS was the runner-up with the AUC of 0.8527 (95%CI: 0.7405-0.9650), followed by KNN and SVDD with AUC of 0.8521 (95%CI: 0.7250-0.9792) and 0.8267 (95%CI: 0.7013-0.9521), respectively. Regarding single-modal features, the MoG model constructed on CFA outperformed those on MRI (AUC of 0.6984, with 95%CI of 0.6551-0.7418), MH&DEM (AUC of 0.5938, with 95%CI of 0.4076-0.7800), and BLO&ApoE (AUC of 0.5920, with 95%CI of 0.5394-0.6446) (see Additional file 1: Table S3).

Additionally, when using the combination of CFA and IMG modalities, all ND models produced the best detection performance with the sensitivity for MCI patients, especially, the MoG presented the highest sensitivity of 69.86%, followed by KNN, SVDD, and KMEANS with sensitivity of 68.61%, 68.57%, and 67.60%, respectively. Hence, CFA features are most discriminative, while IMG markers provide additional

evidence for detecting MCI. Further adding BLO&ApoE and MH&DEM to the combination of CFA and IMG cannot make the model distinguish MCI better and even reduce the sensitivity of MCI. E.g., By adding those features, KNN produced low sensitivity of MCI, dropping from 68.61% to 62.64%, but with relatively stable AUC performance (Table 2). It is worthy of note that models built on MH&DEM and BLO&ApoE modalities achieved the worst performance with respect to the AUC, sensitivity, and specificity metrics.

FMUUH data

Consistent with the experimental results on AIBL, those on FMUUH data shows that (Table 3), with the CFA only, MoG again produced the highest average AUC of 0.9443 (95%CI: 0.9037-0.9849) (Table 3, bold), and the KMEANS was the runner-up with the AUC of 0.9330 (95%CI: 0.8842-0.9818), followed by KNN and SVDD with AUC of 0.9299 (95%CI: 0.8820-0.9777) and 0.8386 (95%CI: 0.7435-0.9334), respectively.

Table 3 The performance comparison of the ND models constructed by KNN, MoG, KMEANS, and SVDD based on the FMUUH data.

Modality combinations	ND algorithms	False rejection rate: 0~1	False rejection rate = 0.1	
		AUC (95%CI)	Specificity HC	Sensitivity AD
CFA (ADCS-ADL, NPI, MMSE,)& DEM(Age,	KNN	0.9022 (0.8587,0.9457)	89.13%	88.27%
	MoG	0.8989 (0.8383,0.9695)	91.01%	89.74%
	KMEANS	0.8869 (0.8332,0.9406)	88.45%	85.58%

Gender, Education)	SVDD	0.7497 (0.7120,0.7874)	86.89%	79.14%
CFA (ADCS-ADL, NPI, MMSE)	KNN	0.9299 (0.8820,0.9777)	93.28%	82.15%
	MoG	**0.9443 (0.9037,0.9849)**	**90.92%**	**89.09%**
	KMEANS	0.9330 (0.8842,0.9818)	92.56%	86.31%
	SVDD	0.8386 (0.7435,0.9334)	84.75%	82.42%
DEM(Age, Gender, Education)	KNN	0.6921 (0.6654,0.7188)	66.67%	60.05%
	MoG	0.7223 (0.6335,0.8111)	72.16%	63.15%
	KMEANS	0.7347 (0.6551,0.8143)	72.05%	61.58%
	SVDD	0.5128 (0.3778,0.6478)	71.01%	63.56%

Decision boundary constructed on HC only

Fig.3 illustrates the decision boundary produced by each novelty detector. Randomly, the 1st fold data were selected from both AIBL and FMUUH datasets. For data visualization, the two most important features were used according to the feature selection results. Therefore, the MMSE and ADCS-ADL features were selected from the FMUUH data, while MMSE and LMDR were used for the AIBL data. All selected data have been scaled between 0 and 1. To quantify the detection performance of the four ND methods on the testing data, Table 4 lists the mean of 5-fold sensitivity and specificity performance, representing the proportion of Non-HC (i.e. MCI & AD) data lying outside the boundary and the proportion of HC data lying inside the boundary.

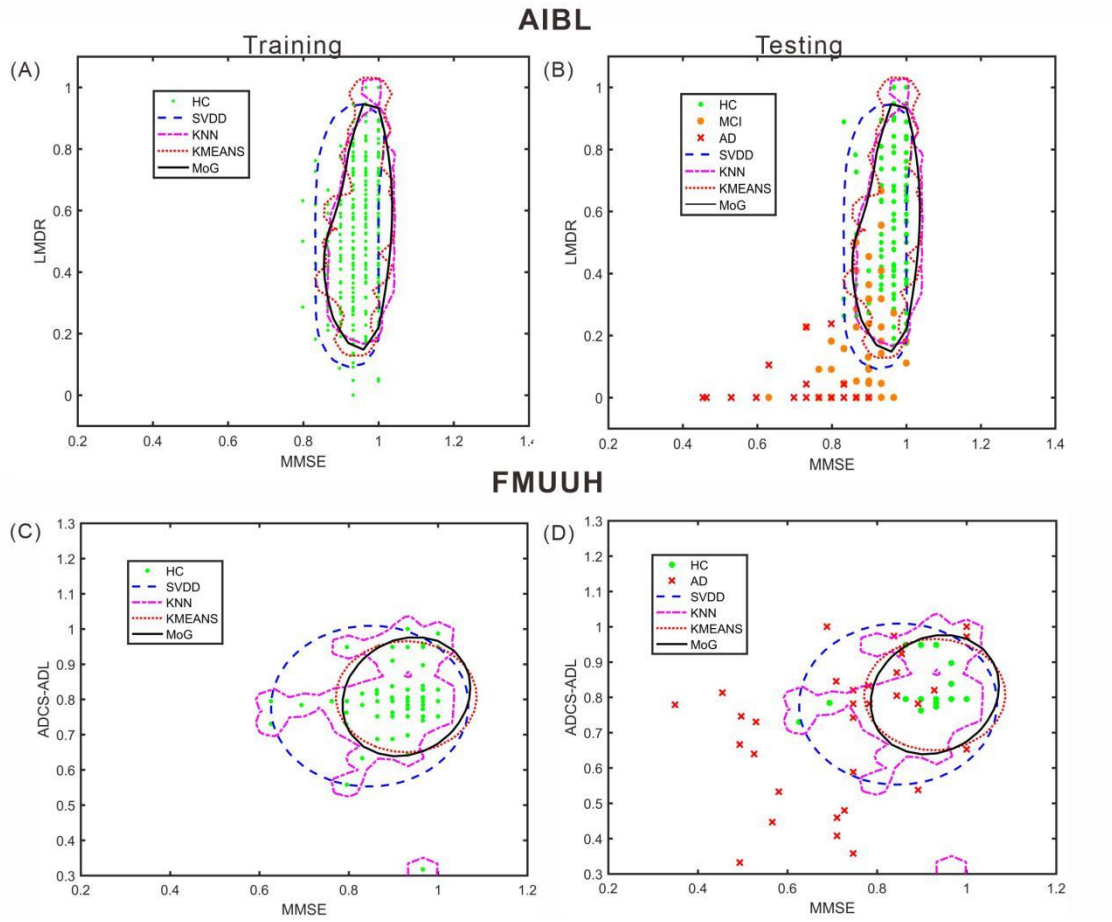

Fig.3 The decision boundaries produced by the four novelty detection methods. (A) & (C). Boundaries trained on HC only in terms of the AIBL and FMUUH data respectively. (B) & (D). The trained boundaries were used to test the testing data including both HC and Non-HC with respect to the AIBL and FMUUH data respectively. The boundaries produced by the SVDD, KNN, KMEANS and MoG methods are represented by blue dashed, magenta dotted, cyan dotted and black solid lines, respectively. Green dots, orange dots, and red crosses indicate HC, MCI, and AD. The training threshold (i.e. false rejection rate) is set to 0.1.

Table 4 The quantitative evaluation of the visualised decision boundaries constructed by four ND methods (KNN, MoG, KMEANS, and SVDD) on both AIBL and FMUUH data sets.

	AIBL			**FMUUH Data**	
	Specificity	Sensitivity		Specificity	Sensitivity
	HC	MCI	AD	HC	AD
KNN	88.83%	64.88%	98.01%	89.94%	70.59%
MoG	88.72%	67.95%	98.01%	87.99%	76.96%

KMEANS	86.13%	60.32%	99.69%	88.27%	74.93%
SVDD	88.21%	51.96%	97.40%	94.15%	54.41%

Sensitivity for MCI and AD and specificity for HC were calculated based on 5-fold CV assessment.

The analysis of AIBL data

All trained boundaries shown in Fig. 3A can enclose at least 86% HC data, but the MoG produced a tighter and smoother boundary than others. In term of the testing results (Table 4), all methods accurately distinguished AD from HC with high sensitivity (higher than 97%). Especially, the MoG boundary rejected 98.01% AD and accepted 88.72% HC. On the other hand, the sensitivity for MCI performed worse than that for AD. Although boundaries generated by the MoG, KNN, SVDD accepted higher than 88% HC, they only rejected 67.95%, 64.88%, and 51.96% MCI. Hence the MoG is less likely to misdiagnosis for MCI and will be more suitable for early diagnosis or disease warning than other methods. Linking to Fig.3B, some MCI data points (orange dots) lie inside the boundaries, indicating some overlap between HC and MCI. This overlap may because MCI cannot be judged using only two features. Nevertheless, the distance from a data point to the decision boundary can objectively reflect the risk and severity of developing MCI or AD for an individual to a certain extent. From this point of view, the boundary generated by ND methods would be inspiring for solving the problem of clinically unclear diagnostic criteria for MCI. There is a benefit that the ND technique introduces to early diagnosing and prognosing of MCI/AD.

The analysis of FMUUH data

As decentralized data distribution, multiple boundaries were generated by the KNN, and one loose boundary was produced by the MoG, KMEANS and SVDD (Fig. 3C), trying to include at least 88% HC. Some overlap between HC and AD (Fig. 3D) reflected high specificity for HC but low sensitivity for AD (Table 4). E.g., the MoG obtained the lowest specificity of 87.99% for HC, but the highest sensitivity of 76.96% for AD. Therefore, we proposed a Distance to Boundary (DtB) strategy to address this inevitable low sensitivity caused by overlap between HC and non-HC (MCI/AD) and further detect potential non-HC from HC.

The DtB strategy

The idea behind the DtB strategy comes from that the ND decision boundary can objectively reflect the severity of developing AD of individuals, according to the distance of data point to the boundary. The theoretical foundation of the strategy was proposed by calculating the distance of each data point to its nearest point on the boundary in order to quantify the severity of cognitive decline. To describe the strategy, we choose the MoG algorithm, which has a stable boundary, best overall performance and can precisely enclose data, to calculate the DtB values of all testing data. The DtB calculation was carried out on the 5-fold CV assessment results. Fig. 4 depicts the box plots of the distance of each category of HC, MCI, and AD data points to the decision

boundary constructed by MoG (Fig. 4A for AIBL and Fig. 4B for FMUUH). We define the sign of distance of inner data to the boundary is negative, while that for outer data is positive. Table 5 lists descriptive statistics of the results in Fig. 4.

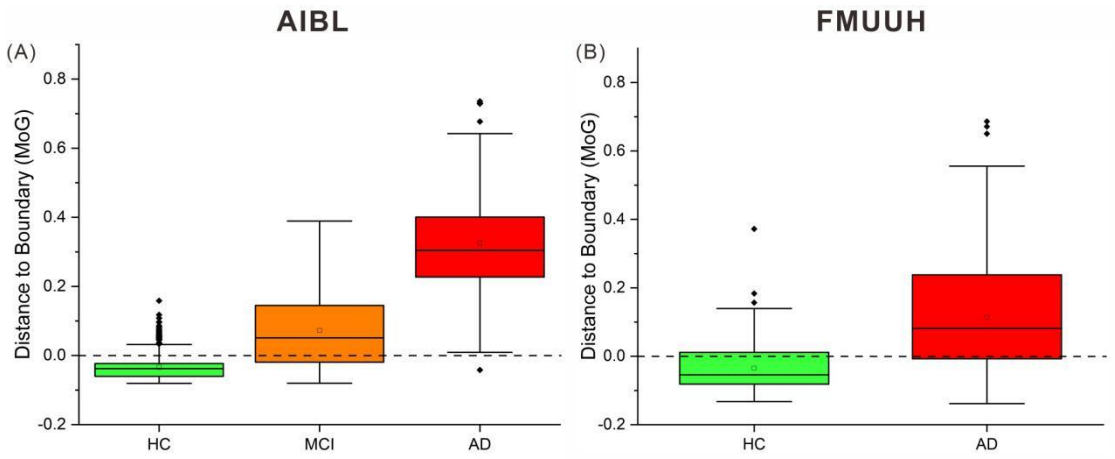

Fig.4 Box plot of the distance between AIBL and FMUUH data points and ND boundary generated by MoG. Points beyond 1.5 times IQR (interquartile range) are considered outliers of the box plot, represented by a solid diamond. The hollow square is the mean value, and the dotted line is the ND boundary location. (A) for the AIBL data and (B) for the FMUUH data.

Table 5 Descriptive statistics of DtB results.

		Mean	Mode	Minimum	Median	Maximum
AIBL	HC	-0.0342	-0.0665	-0.0806	-0.0377	0.1587
	MCI	0.0725	0.2587	-0.0800	0.0511	0.3890
	AD	0.3258	0.3052	-0.0420	0.3045	0.7356
FMUUH	HC	-0.0348	-0.0813	-0.1324	-0.0545	0.3722
	AD	0.1136	0.0813	-0.1384	0.0813	0.6860

Fig. 4A turns out that the 1st and the 3rd quartiles, median and mean, and maximum DtB values of AD are higher than those of MCI, which are in turn higher than those of HC. The length of the boxes of MCI and AD is more than twice that of the HC. The overall spreads are quite different; even some overlaps occur among the three categories. Compared to the main bodies of data for MCI and AD which look symmetric, the boxplot for HC shows some slight bottom-skew since there are more outliers which would affect any calculations of skewness. The potential outliers of HC may risk getting MCI. Similarly, the outliers of MCI may be developing AD, and the outliers of AD will be developing more and more severely. Overall the HC, MCI, and AD categories do vary with the DtB values. Fig.4A could interestingly visualize our proposed DtB strategy, i.e. for those data points lying outside the boundary, the farther from the boundary, the more severely the representing patients may develop AD; whereas for the inner data points, the nearer to the boundary, the more risk of cognitive decline they reflect.

Basically, the DtB box plot of FMUUH data shows a similar trend to that of AIBL. The AD box is longer than HC's, as well as presenting higher 1st and 3rd quartiles, median, mean, and maximum DtB values than HC (Fig. 4B). The difference is that the AD box of FMUUH data is slight across the ND boundary (the horizontal dotted line), which is similar to the MCI box of AIBL. Although an AD data point may be misdiagnosed as HC due to its lying inside the boundary, we can still detect its risk of developing AD according to its DtB value. Note that, however, different features used in Fig. 3 (A-B) and (C-D) and different population cohorts from Australian and China may cause the

difference between Fig. 4 (A) and (B). Meantime, Fig. 4B hints some MCI might be included in the AD patients, which also reflects an urgent need of data integration of various data resources based on the Fujian Medical University Union Hospital.

Table 5 reflects that the minimum DtB values of HC and MCI in AIBL are very close (i.e. -0.0806 and -0.0800). Again this may due to overlap of HC and MCI to some extent and the lower boundary dimension or the ambiguity of 'cutoff' scores for determining MCI [37]. In a higher-dimensional feature space, the DtB strategy can integrate more different assessment indicators and criteria to describe more precisely the severity of MCI patients. Similarly, close minimum DtB values of HC and AD in FMUUH reflects the beforementioned hint.

Discussion

This study first employed ND techniques to detect potential MCI/AD from HC populations. Four representative and easily interpretable ND methods were applied on the AIBL and local FMUUH datasets to construct an optimal and closed decision boundary tightly surrounding the given HC data only. The surface area of the decision boundary should be minimized to reduce the chance of accepting MCI/AD data. Once getting the boundary, the model will be able to classify if unseen data reflect healthy or non-healthy depending on it lying inside or outside the boundary. Inner data that

are close to the boundary have a high risk of developing MCI even they may be currently detected as HC. For those outer data, the closer to the boundary the milder the cognition declines (e.g. very mild or mild cognitive impairment) that the data represent; while the farther from the boundary, the severer the cognition declines (e.g. moderate or severe AD). This was quantified by our proposed DtB strategy. Data features involved in the study are multi-modal. The ND models built on different modalities and their combinations were evaluated using three comprehensive metrics.

In terms of the AIBL data, ND methods produced comparable detection performance when only a small subset of the data was used, and this subset mainly composed of easily accessible CFA (LMDR, LMIR, MMSE). The KNN, MoG and KMEANS achieved the highest performance when only the CFA was used to train the models (AUC = 0.8591, 0.8878 and 0.8687 respectively) (Table 1). This result once again highlights the CFA could be a key factor of diagnosing AD used in clinical practices [38]. The extensive experimental results turn out that using the CFA only could always reflect the best global detection performance (the AUC metric), while combining IMG would perform the best performance in terms of sensitivity for detecting MCI from HC. For example, the KNN method produced a higher sensitive for MCI (68.61%) based on a combination of CFA and IMG in comparison to using CFA only (59.87%). (Table 2). The MoG method has the same trend. It is worthy of note that the highest detection performance for MCI (sensitivity of 69.86%) was achieved by the MoG-based ND model trained with CFA and IMG (Table 1). Therefore, the model trained with the combination of IMG and

CFA modalities outperformed that constructed with CFA only in the task of early identification of MCI patients.

Interestingly, feature extraction does not significantly improve the detection results of KNN and even reduces it. For example, after feature selection, the KNN model performance is AUC=0.6948 (0.5944-0.7953) using the combination of IMG, MH&DEM and BLO&ApoE. However, when feature selections are not applied, its AUC rises to 0.8113, 95%CI=(0.7402-0.8825) (Additional file 1:Table S3). This difference indicates that the KNN algorithm has good performance in high dimensional feature spaces and may have a strong ability to resist noise information.

The FMUUH data, derived from the populations with different geographical locations and races in comparison to the AIBL, were used to verify the applicability of the ND model. Similarly, MoG produced the best performance with the highest average AUC of 0.9443 (95% CI: 0.9037-0.9849) when using the CFA only, as well as sensitivity of 89.09% for AD and specificity of 90.92% for HC (Table 3 bold). The proposed DtB strategy suggested that policymakers should positively focus on an urgent need for data integration in the area.

This study has several limitations worth noting, which in turn bring future extensions and improvements. First, we ignored missing data. An approach for missing data

imputation is currently being developed, which will be incorporated into the system later. Second, we only selected the top 10 features from the AIBL data that are significant to the CDR, using three different univariate filters to rank the feature importance. However, such filtering approaches may lead to the loss of relevant features that are meaningless by themselves, but when considered together, can improve model performance. To overcome this, we have previously applied wrapper methods to evaluate the importance of specific feature sets [39]. Work is currently in progress to improve our ND model by developing a wrapper that can obtain a subset of better-integrated data from different modalities. Third, to get a large data size, we integrated the multi-modal AIBL data collected at different time points together. Repeat visitors who participated in the AIBL study were considered as different visitors. However, some modalities such as medical history, ApoE genotype, and gender are not time-evolved. Those features may be the reason for poor performance generated when the models were trained by the two modalities of MH&DEM and BLO&ApoE. We are further investigating the ND technique on a larger-sized data, e.g. ADNI (The Alzheimer's Disease Neuroimaging Initiative) [40] or NACC (The National Alzheimer's Coordinating Center) [41] data, and working on integrating more FMUUH data from the local hospital. Furthermore, the current study and our previous development [39] have provided a solid foundation for the next phase extension of developing a CDSS using the ND technique. Finally, Ding et al. [42] have proposed a level set method based ND approach (LSM-ND), namely level set boundary description (LSBD). Being superior to the traditional methods (e.g. based on probability, distance, clustering,

statistics, and support vector machines), the LSBD introduced some interesting properties for boundary construction, such as nonlinear problem addressable without using a kernel trick, non-parametric, dynamically time-evolved to better fit the data distribution, boundary shape easily manageable, and straightforward implemented in the given data space, et al.. Therefore, based on the current work, we will deeply investigate the LSM-ND approach for early detecting MCI and AD from HC populations.

Conclusion

This study first built the novelty detectors based on different Alzheimer's datasets (AIBL and FMUUH) using four representative and easily interpretable ND algorithms. The intrinsic pattern behind AD was investigated in different cohorts study, by comparing the performance of models trained not only by different modalities but also the different combinations of modality types. We found that the best overall performance can be obtained when only CFA features are used. Hence, applying some non-invasive and easily accessible features can significantly detect cognitive decline in an early stage. Moreover, the highest detection sensitivity of MCI can be obtained by a combination of CFA and IMG. More importantly, the insight of the study was presented by a proposed DtB strategy via illustrating and quantifying the decision boundary produced by each novelty detector using two (for visualization purpose) of the most significant CFA features from each dataset. The strategy can objectively

reflect the severity of developing AD of individuals. These results would help inform future guidelines for the development of a clinical decision-making support system aiming at an early diagnosis and prognosis of MCI/AD.

Additional file 1: Fig.S1 Demographic features distribution of the FMUUH data. **Table S1.** Demographic distribution of AIBL data. **Table S2.** The performance of the ND models using all possible modality combinations after feature selection. **Table S3.** The performance of ND models using all possible modality combinations without feature selection.

Abbreviations

AD: Alzheimer's disease; ADI: Alzheimer's Disease International; ND: Novelty detection; AI+CDSS: AI-enabled clinical diagnosis support systems; FMUUH: Fujian Medical University Union Hospital; SVDD: support vector data description; KNN: K Nearest Neighbor； MoG: Mixture of Gaussian; CSF: cerebrospinal fluid biomarkers; CFA: cognitive and functional assessments; MRI: magnetic resonance imaging; PET: Positron emission tomography; CDR: The Clinical Dementia Rating; MMSE: Mini-Mental State Examination; LMIR/LMDR: logical memory immediate/delayed recall assessments; ADCS-ADL: Alzheimer's Disease Cooperative Study - Activities of Daily Living; NPI: The Neuropsychiatric Inventory; CES-D: the Center for Epidemiologic Studies Depression Scale; SF-12: The Short Form 12; ECog-SP&Ecog-Pt: Everyday Cognition; ROC: the receiver operating characteristic; AUC: area under the ROC curve; MH&DEM: medical history and demographics; BLO&ApoE: blood test and two alleles apolipoprotein genotyp; IGR: information gain ratio; HC: healthy controls; MCI: mild cognitive impairment; AXT117: hormone; BAT126: vitamin B12; HMT3: red blood cell; HMT7: white blood cell; HMT13: platelets; HMT40: haemoglobin; HMT100: mean corpuscular haemoglobin; HMT102: mean corpuscular haemoglobin concentration; RCT6: urea nitrogen; RCT11: serum glucose; RCT20: cholesterol (high performance); RCT392: creatinine(rate blanked).

Acknowledgements

Not applicable.

Authors' contributions

J.M. and X.D. performed the data analysis and data modelling, and prepared the manuscript. Q.Y. provided the clinical data and consultation on the data. H.Y. and X.D. managed the project process.

S.L. provided the preliminary work of the project and helped prepare the manuscript. W.G. helped prepare the manuscript. Y.W. and J.X. helped provided the clinical data.

Funding

This project was supported by the Alzheimer's Research UK NI Networking (X.D., H.Y., Q.Y.), the National Natural Science Foundation of China under grant no.81870995, Global Challenges Research Fund Networking (X.D., H.Y., Q.Y., M.B.), the National Key Basic Research Program of China under grant no.2015CB352006, the scientific research innovation team construction program of Fujian Normal University under grant no. IRTL1702, Special Funds of the Central Government Guiding Local Science and Technology Development under grant no. 2017L3009, the Natural Science Foundation of Fujian Province under grant no.2018J01814 and the Youth Natural Science Research Programme, Fujian, China (JZ160425).

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

The study was approved by Fujian Medical University Union Hospital and AIBL, which waived the requirement for informed consent due to the nature of the study. All data were anonymized to maintain confidentiality.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

[1]. Feigin, V.L., et al., The global burden of neurological disorders: translating evidence into policy. The Lancet Neurology, 2020. 19(3): p. 255-265.

[2]. Simonetti, A., et al., Neuropsychiatric symptoms in elderly with dementia during COVID-19

pandemic: definition, treatment, and future directions. Frontiers in psychiatry, 2020. 11.

[3]. International, A.S.D., World Alzheimer report 2019: attitudes to dementia. 2019, Alzheimer’s Disease International London, UK.

[4]. Zhu, Y., et al., Prevalence of dementia in the People’s Republic of China from 1985 to 2015: a systematic review and meta-regression analysis. BMC public health, 2019. 19(1): p. 578.

[5]. Xu, J., et al., The economic burden of dementia in China, 1990–2030: implications for health policy. Bulletin of the World Health Organization, 2017. 95(1): p. 18.

[6]. Aljaaf, A.J., et al. Toward an optimal use of artificial intelligence techniques within a clinical decision support system. in 2015 Science and Information Conference (SAI). 2015: IEEE.

[7]. Hall, W. and J. Pesenti, Growing the artificial intelligence industry in the UK. Department for Digital, Culture, Media & Sport and Department for Business, Energy & Industrial Strategy. Part of the Industrial Strategy UK and the Commonwealth, 2017.

[8]. Van Hartskamp, M., et al., Artificial intelligence in clinical health care applications. Interactive Journal of Medical Research, 2019. 8(2): p. e12100.

[9]. Pascual, D.G., Artificial intelligence tools: decision support systems in condition monitoring and diagnosis. 2015: Crc Press.

[10]. Ding, X., et al., An experimental evaluation of novelty detection methods. Neurocomputing, 2014. 135: p. 313-327.

[11]. Deep, S., et al., A survey on anomalous behavior detection for elderly care using dense-sensing networks. IEEE Communications Surveys & Tutorials, 2019. 22(1): p. 352-370.

[12]. Mohammadian Rad, N., et al., Novelty detection using deep normative modeling for imu-based abnormal movement monitoring in parkinson’s disease and autism spectrum disorders. Sensors, 2018. 18(10): p. 3533.

[13]. Nugroho, L.E., L. Lazuardi and A.S. Prabuwono. Detection of Anomalous Vital Sign of Elderly Using Hybrid K-Means Clustering and Isolation Forest. in TENCON 2018-2018 IEEE Region 10 Conference. 2018: IEEE.

[14]. Vasighizaker, A., A. Sharma and A. Dehzangi, A novel one-class classification approach to accurately predict disease-gene association in acute myeloid leukemia cancer. PloS one, 2019. 14(12): p. e0226115.

[15]. Chang, R.M., R.J. Kauffman and Y. Kwon, Understanding the paradigm shift to computational social science in the presence of big data. Decision Support Systems, 2014. 63(0): p. 67-80.

[16]. Ellis, K.A., et al., The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. International psychogeriatrics, 2009. 21(4): p. 672-687.

[17]. Escobar, J.I., et al., Use of the Mini-Mental State Examination (MMSE) in a community population of mixed ethnicity. Cultural and linguistic artifacts. J Nerv Ment Dis, 1986. 174(10): p. 607-14.

[18]. Ding, X., et al., A hybrid computational approach for efficient Alzheimer's disease classification based on heterogeneous data. Sci Rep, 2018. 8(1): p. 9774.

[19]. Galasko, D., et al., ADCS Prevention Instrument Project: assessment of instrumental activities of daily living for community-dwelling elderly individuals in dementia prevention clinical trials. Alzheimer Dis Assoc Disord, 2006. 20(4 Suppl 3): p. S152-69.

[20]. Cummings, J.L., The Neuropsychiatric Inventory: assessing psychopathology in dementia patients. Neurology, 1997. 48(5 Suppl 6): p. S10-6.

[21]. Karegowda, A.G., A.S. Manjunath and M.A. Jayaram, Comparative study of attribute selection using gain ratio and correlation based feature selection. International Journal of Information Technology and Knowledge Management, 2010. 2(2): p. 271-277.

[22]. Biesiada, J. and W. Duch, Feature selection for high-dimensional data—a Pearson redundancy based filter, in Computer recognition systems 2. 2007, Springer. p. 242-249.

[23]. Jin, X., et al. Machine learning techniques and chi-square feature selection for cancer classification using SAGE gene expression profiles. in International Workshop on Data Mining for Biomedical Applications. 2006: Springer.

[24]. Lin, S., J. Ding and J. Zhou. Rank aggregation of putative microrna targets with cross-entropy monte carlo methods. in Preprint, presented at the IBC 2006 conference, Montral. 2006.

[25]. Schölkopf, B., et al., Estimating the support of a high-dimensional distribution. Neural computation, 2001. 13(7): p. 1443-1471.

[26]. Bishop, C.M., Novelty detection and neural network validation. IEE Proceedings-Vision, Image and Signal processing, 1994. 141(4): p. 217-222.

[27]. Tax, D.M. and R.P. Duin, Support vector data description. Machine learning, 2004. 54(1): p. 45-66.

[28]. Ding, X., et al., An experimental evaluation of novelty detection methods. Neurocomputing, 2014. 135: p. 313-327.

[29]. Pimentel, M.A., et al., A review of novelty detection. Signal Processing, 2014. 99: p. 215-249.

[30]. Markou, M. and S. Singh, Novelty detection: a review—part 1: statistical approaches. Signal processing, 2003. 83(12): p. 2481-2497.

[31]. Faria, E.R., et al., Novelty detection in data streams. Artificial Intelligence Review, 2016. 45(2): p. 235-269.

[32]. Hautamaki, V., I. Karkkainen and P. Franti, Outlier detection using k-nearest neighbour graph, in Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.

Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004. 2004. p.

430-433 Vol.3.

[33]. Bishop, C.M., Pattern recognition and machine learning. 2006: springer.

[34]. Varshney, K.R. and A.S. Willsky, Classification using geometric level sets. The Journal of Machine Learning Research, 2010. 11: p. 491-516.

[35]. Lazzaretti, A.E. and D.M. Tax. An adaptive radial basis function kernel for support vector data description. in International Workshop on Similarity-Based Pattern Recognition. 2015: Springer.

[36]. Fawcett, T., An introduction to ROC analysis. Pattern recognition letters, 2006. 27(8): p. 861-874.

[37]. Pandya, S.Y., et al., Does mild cognitive impairment always lead to dementia? A review. Journal of the neurological sciences, 2016. 369: p. 57-62.

[38]. Grober, E., et al., Identifying memory impairment and early dementia in primary care. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 2017. 6: p. 188-195.

[39]. Bucholc, M., et al., A practical computerized decision support system for predicting the severity of Alzheimer's disease of an individual. Expert Systems with Applications, 2019. 130: p. 157-171.

[40]. Mueller, S.G., et al., The Alzheimer's disease neuroimaging initiative. Neuroimaging Clinics, 2005. 15(4): p. 869-877.

[41]. Beekly, D.L., et al., NIA Alzheimer’s disease centers. The national Alzheimer’s coordinating center (NACC) database: the uniform data set. Alzheimer Dis Assoc Disord, 2007. 21(3): p. 249-258.

[42]. Ding, X., et al., Novelty detection using level set methods. IEEE Transactions on Neural Networks and Learning Systems, 2014. 26(3): p. 576-588.

Figures

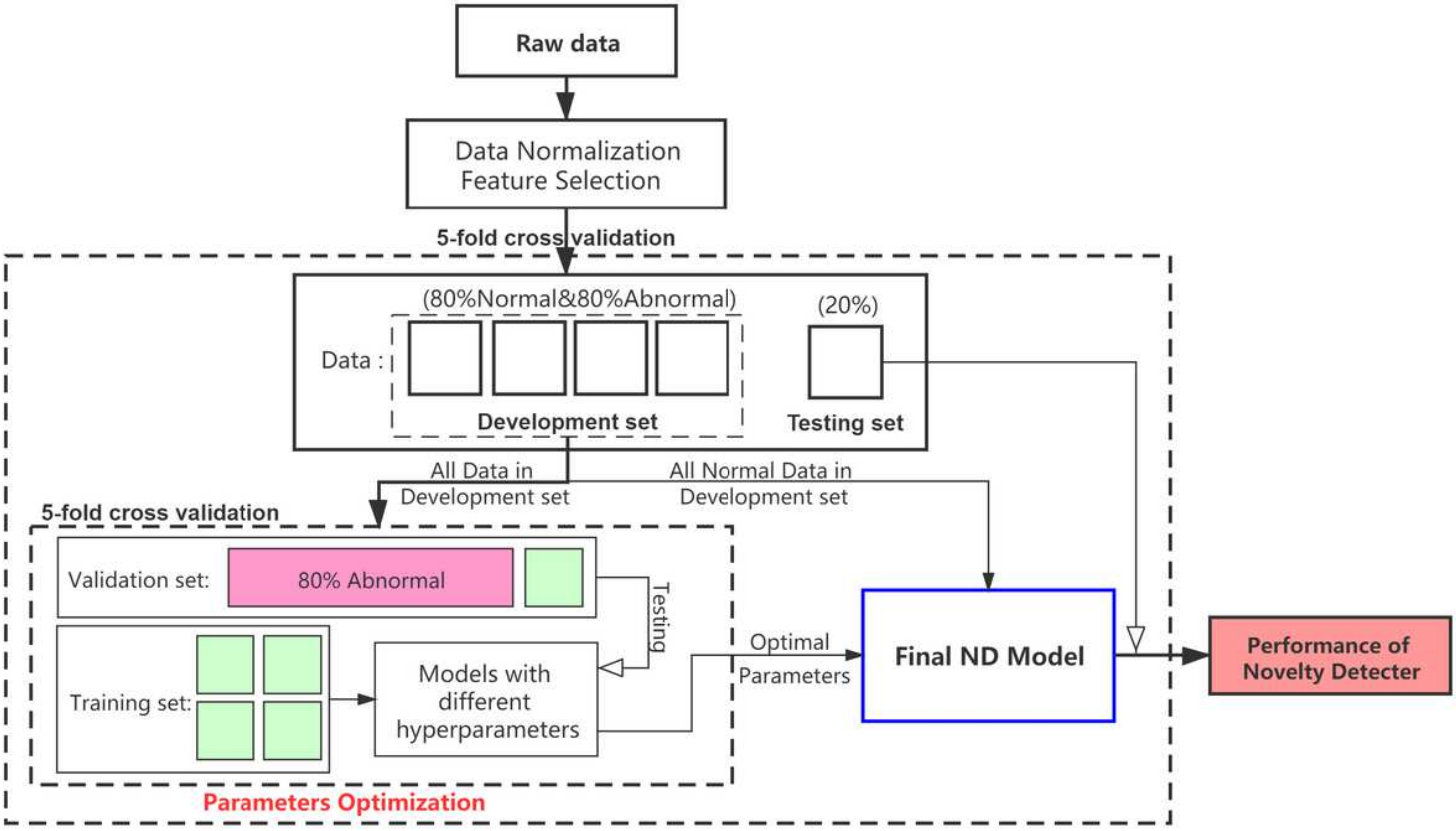

Figure 1

The overall framework of the ND modelling process.

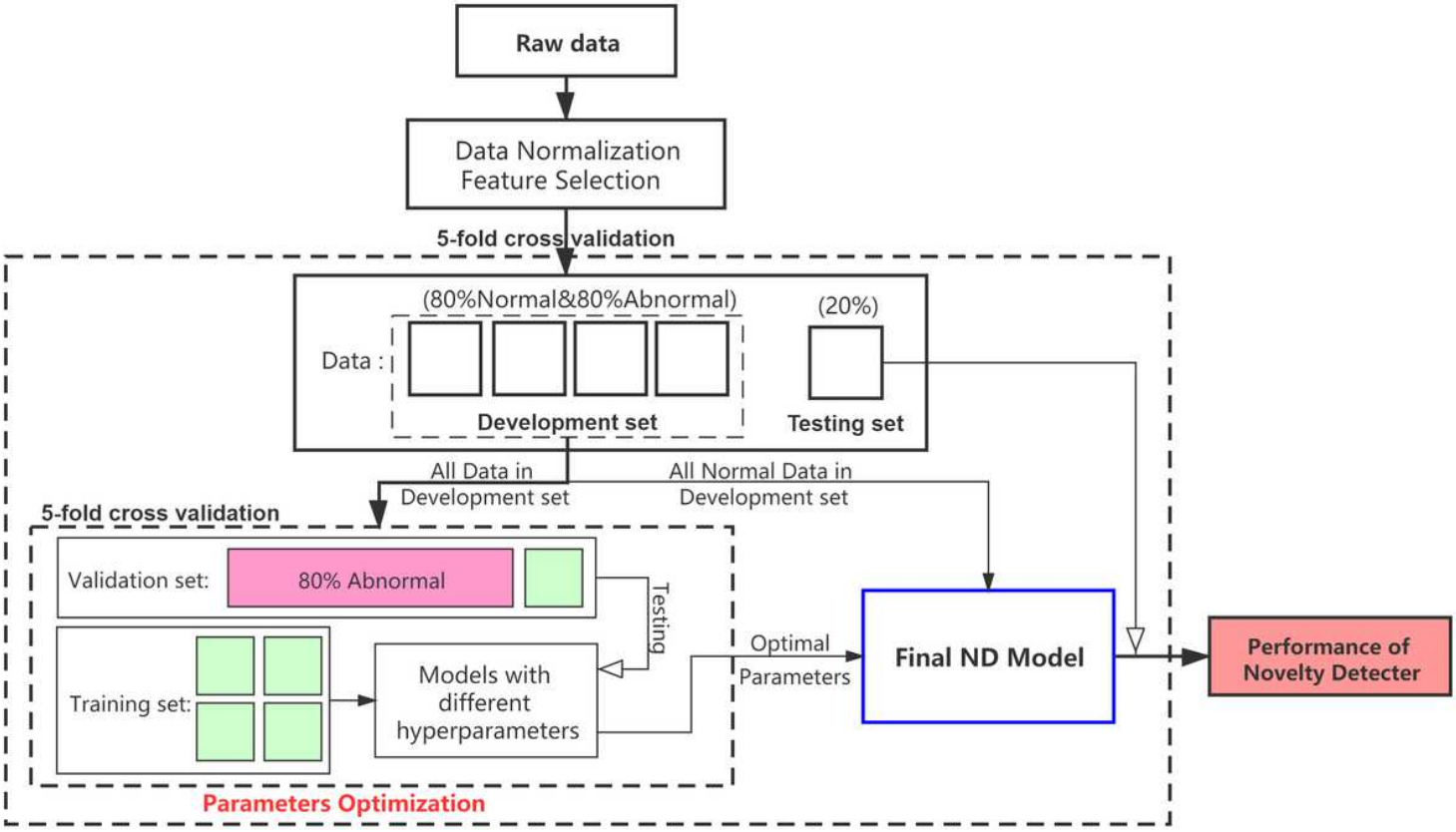

Figure 1

The overall framework of the ND modelling process.

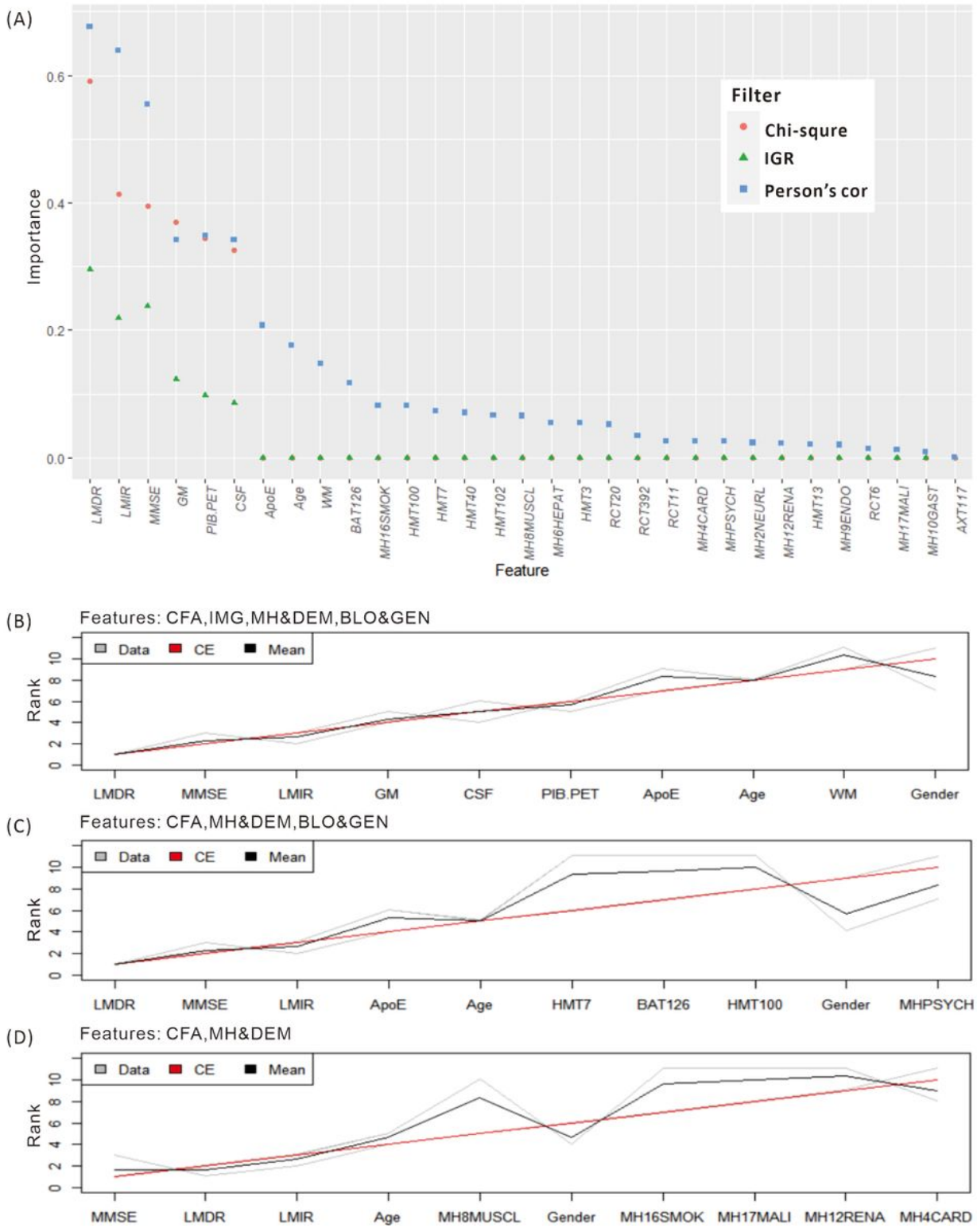

Figure 2

Ranking features importance and aggregating the ranking results with different combinations of modalities. (A). The three filters applied are based on Chi-square value (orange dot), information gain ratio (IGR, green triangle), and Pearson's correlation (blue square). The gray and black lines represent three ranking results and their averages, while the red line reflects the final aggregation result. Some Chi-square values are overlapped by IGR as they all close to zero.

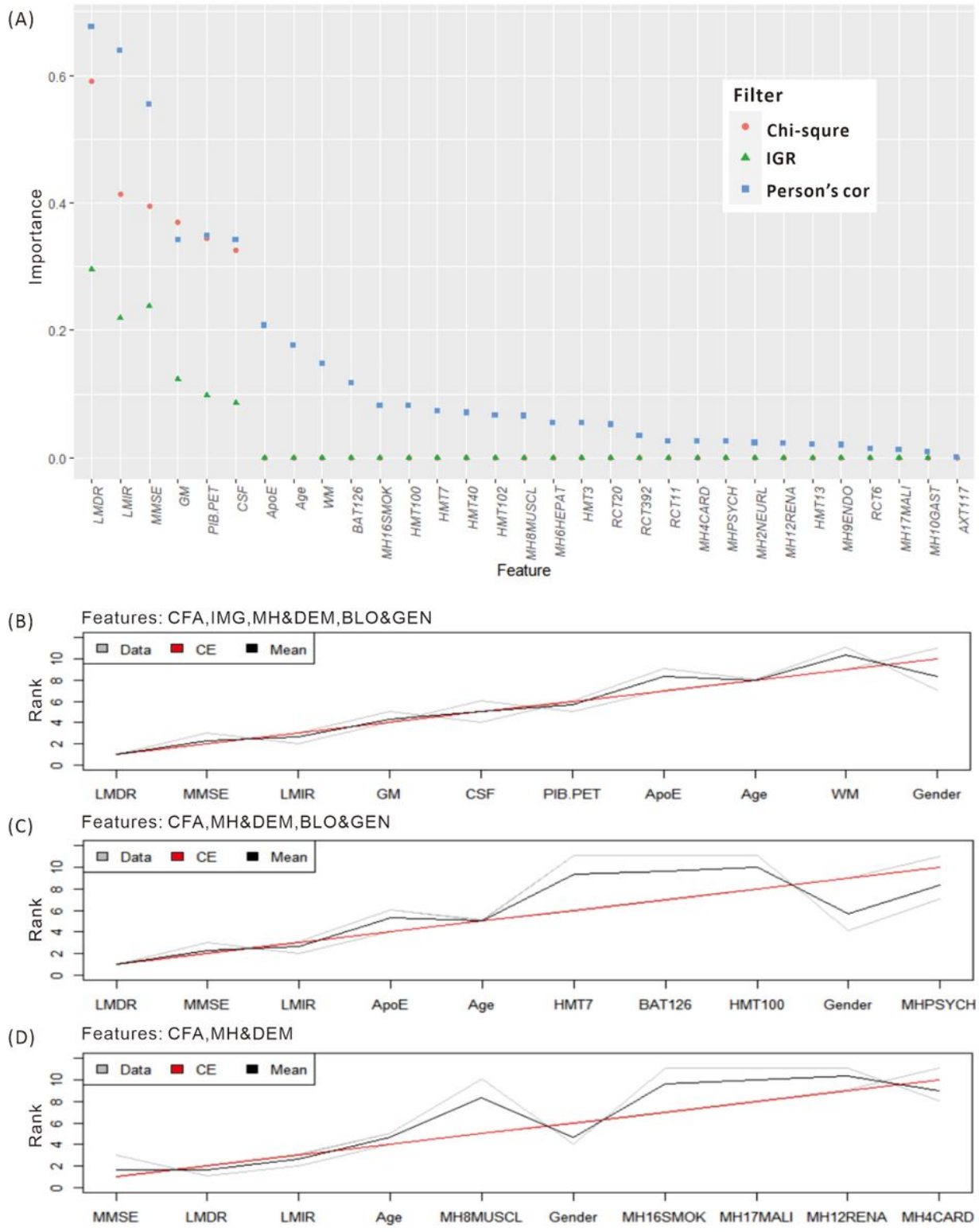

Figure 2

Ranking features importance and aggregating the ranking results with different combinations of modalities. (A). The three filters applied are based on Chi-square value (orange dot), information gain ratio (IGR, green triangle), and Pearson's correlation (blue square). The gray and black lines represent three ranking results and their averages, while the red line reflects the final aggregation result. Some Chi-square values are overlapped by IGR as they all close to zero.

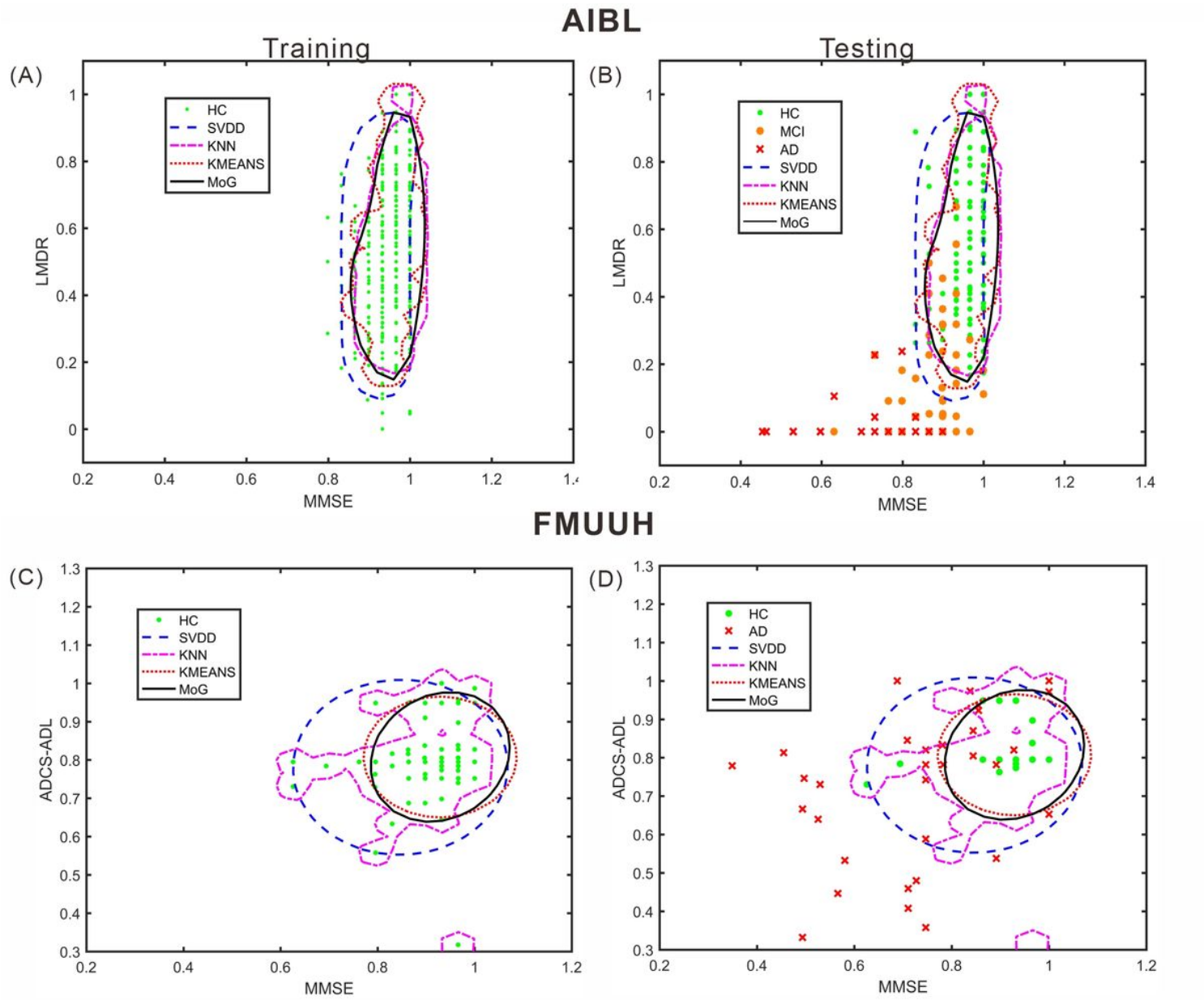

Figure 3

The decision boundaries produced by the four novelty detection methods. (A) & (C). Boundaries trained on HC only in terms of the AIBL and FMUUH data respectively. (B) & (D). The trained boundaries were used to test the testing data including both HC and Non-HC with respect to the AIBL and FMUUH data respectively. The boundaries produced by the SVDD, KNN, KMEANS and MoG methods are represented by blue dashed, magenta dotted, cyan dotted and black solid lines, respectively. Green dots, orange dots, and red crosses indicate HC, MCI, and AD. The training threshold (i.e. false rejection rate) is set to 0.1.

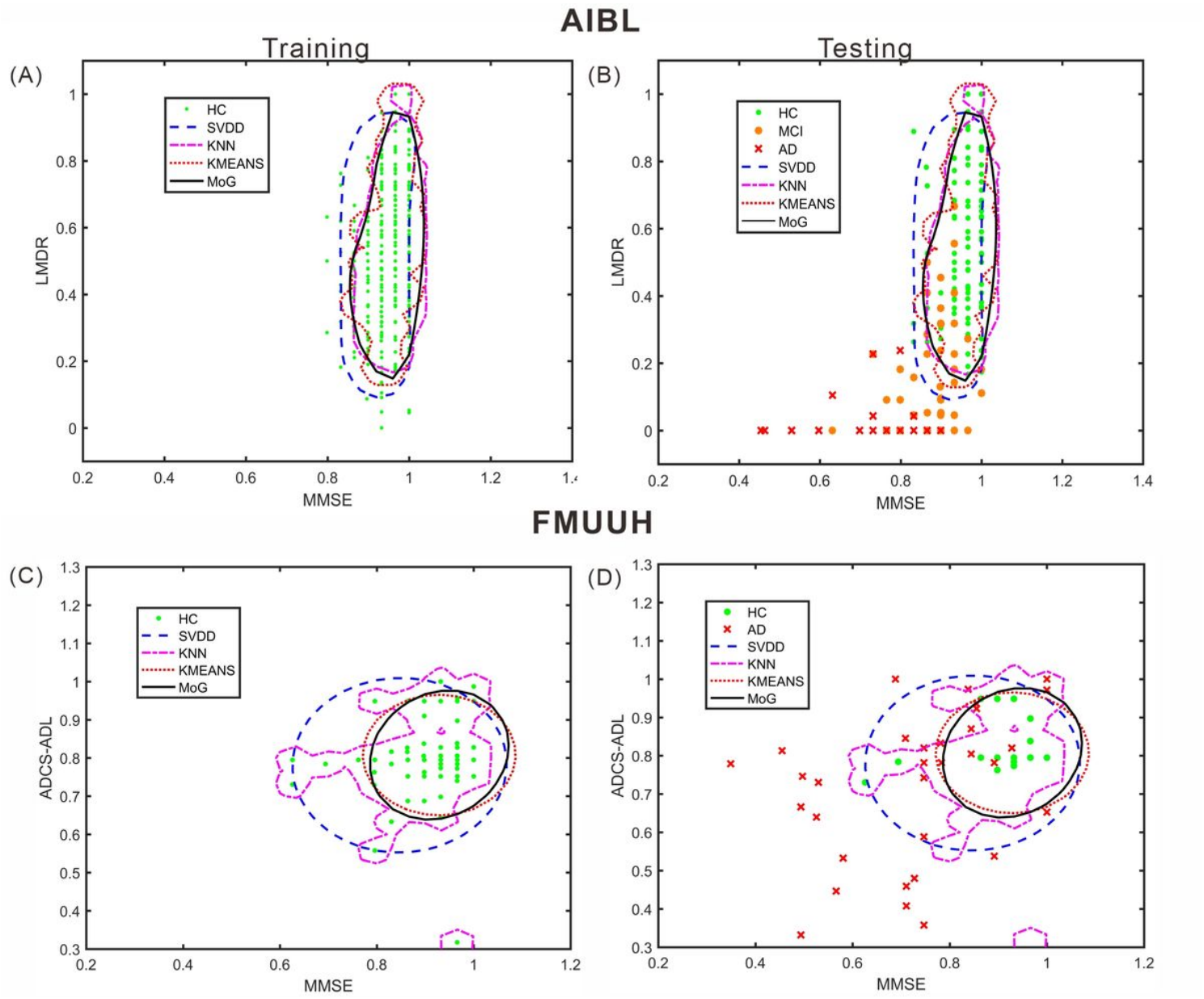

Figure 3

The decision boundaries produced by the four novelty detection methods. (A) & (C). Boundaries trained on HC only in terms of the AIBL and FMUUH data respectively. (B) & (D). The trained boundaries were used to test the testing data including both HC and Non-HC with respect to the AIBL and FMUUH data respectively. The boundaries produced by the SVDD, KNN, KMEANS and MoG methods are represented by blue dashed, magenta dotted, cyan dotted and black solid lines, respectively. Green dots, orange dots, and red crosses indicate HC, MCI, and AD. The training threshold (i.e. false rejection rate) is set to 0.1.

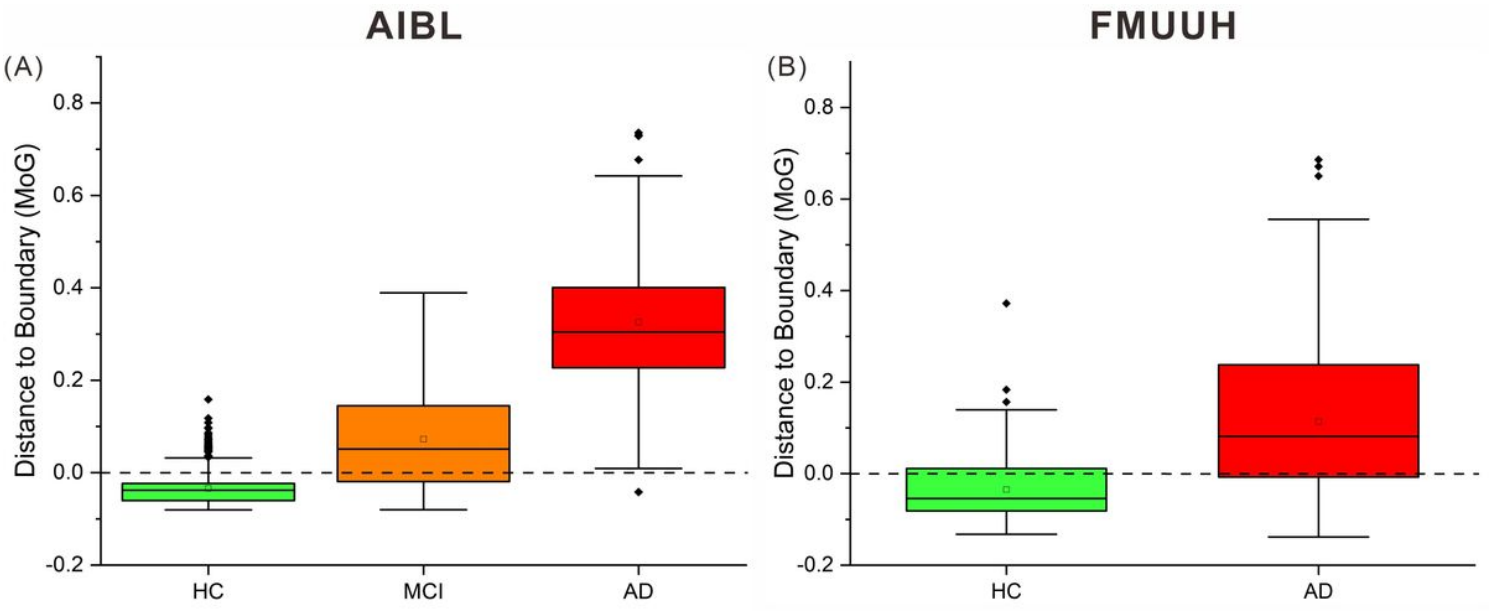

Figure 4

Box plot of the distance between AIBL and FMUUH data points and ND boundary generated by MoG. Points beyond 1.5 times IQR (interquartile range) are considered outliers of the box plot, represented by a solid diamond. The hollow square is the mean value, and the dotted line is the ND boundary location. (A) for the AIBL data and (B) for the FMUUH data.

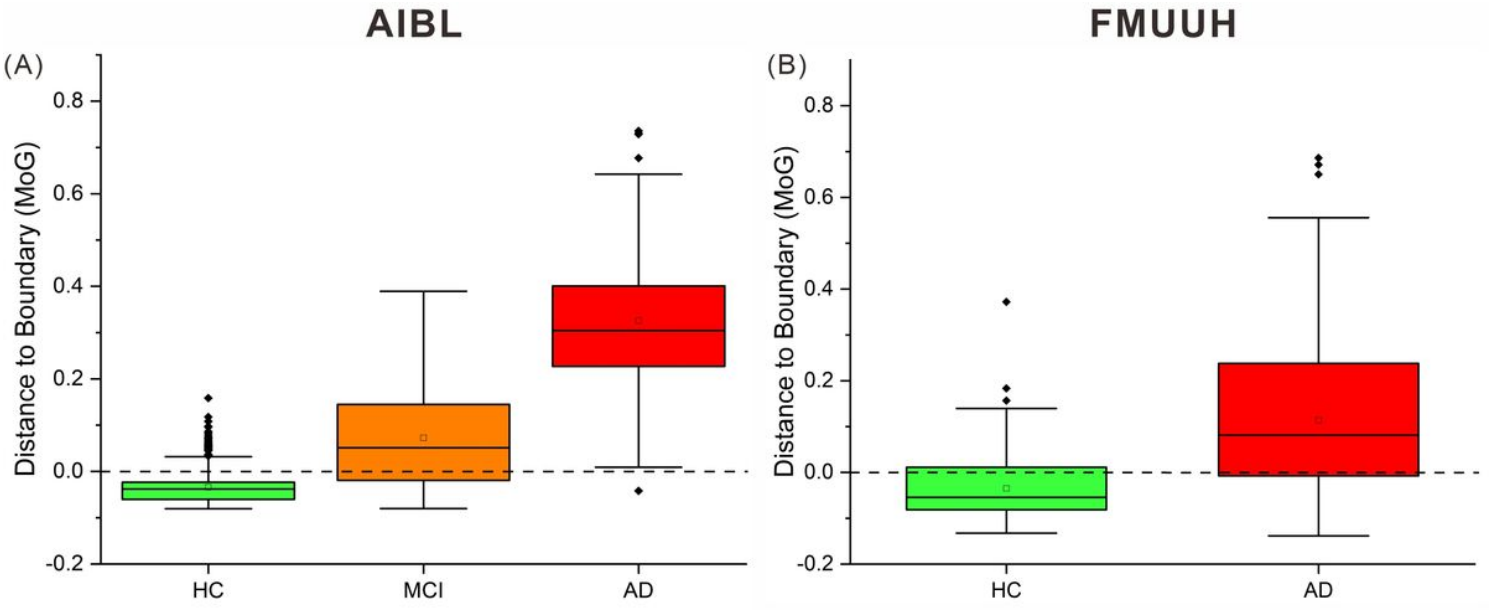

Figure 4

Box plot of the distance between AIBL and FMUUH data points and ND boundary generated by MoG. Points beyond 1.5 times IQR (interquartile range) are considered outliers of the box plot, represented by a solid diamond. The hollow square is the mean value, and the dotted line is the ND boundary location. (A) for the AIBL data and (B) for the FMUUH data.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.doc
- Fig.S1DemographicfeaturesdistributionoftheFMUUHdata.png
- TableS2.xls
- TableS3.xls