Computers & Industrial Engineering 197 (2024) 110625

Contents lists available at ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie

A multimodal learning machine framework for Alzheimer's disease diagnosis based on neuropsychological and neuroimaging data

Meiwei Zhang [a], Qiushi Cui [a,*], Yang Lü [b], Weihua Yu [c], Wenyuan Li [a]

[a] College of Electrical Engineering, Chongqing University, Chongqing, 400030, China
[b] Department of Geriatrics, The First Affiliated Hospital of Chongqing Medical University, Chongqing, 400016, China
[c] Institute of Neuroscience, Chongqing Medical University, Chongqing, 400016, China

ARTICLE INFO

Dataset link: Original embedding pkl (sagittal plane, coronal plane, transverse profile, etc.) (Original data)

Keywords:
Alzheimer's disease
Multimodal deep learning
Extreme learning machine
Ensemble
Neuropsychological tests

ABSTRACT

Alzheimer's disease (AD) is the most prevalent form of dementia, with no current cure. Early screening and intervention are vital. In multimodal AD data, besides neuroimaging dimensions, neuropsychological tests based on cognitive domains also provide the clinical information for diagnosing AD. However, previous multimodal methods often fuse these neuropsychological test scores with other data, losing the rich clinical details inherent in each test, including videos, speech, images, and text. To address this, we propose a novel framework with an entropy-based polynomial dimension expansion function that restores this critical information by accurately calculating the optimal polynomial degree. Additionally, the proposed framework offers a series of cognitive-based Extreme Learning Machine (ELM) models to better utilize the detailed clinical insights from neuropsychological tests, reducing diagnostic redundancy and noise. Finally, we design a boosting ensemble strategy that combines diagnostic models from various dimensions and cognitive domains, automatically optimizing weights to enhance diagnostic accuracy. Tested on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, our approach achieves over 98% accuracy and F1 scores, with no observed bias between mild cognitive impairment (MCI) and AD groups. Therefore, our framework can offer clinicians more logical recommendations for diagnosing and managing the disease.

1. Introduction

Alzheimer's disease (AD) is a complex neurological condition that leads to the global burden of dementia (Gan et al., 2023). It impacts cognitive functions, memory, thought processes, and behavior, characterized by brain plaques and protein tangles (Zhang, Shen, et al., 2023). These disrupt neuronal signaling, leading to cognitive decline and functional impairments, beginning with mild memory loss and gradually progressing to severe dementia and even death. AD leads to memory impairment, language difficulties, confusion, and behavioral changes (Ribeiro et al., 2023), significantly affecting daily life. Despite ongoing research, a cure remains elusive (Kandimalla, Saeed, Tyagi, Gupta, & Aqil, 2023), making AD a major public health challenge with up to 50 million cases globally (Gustavsson et al., 2023).

The identification of cognitive impairment at an early stage poses a complex psychometric challenge, primarily due to the gradual onset of symptoms, often mistakenly attributed to normal age-related cognitive decline (McKee, Hoffos, Vecchiarelli, & Tremblay, 2023). Recent studies have highlighted the potential of neuropsychological tests in accurately detecting the Mild Cognitive Impairment (MCI) stage which is the early phase of AD (Chehrehnegar et al., 2020). Neuropsychological tests, which are categorized based on cognitive domains such as memory, executive function, and attention, offer valuable insights into the subtle cognitive changes associated with AD progression. Furthermore, the modalities through which these tests are administered, including speech, text, images, and videos, provide a multidimensional view of cognitive functioning (Malek-Ahmadi, Davis, Belden, & Sabbagh, 2014). Although these tests contain valuable multimodal information, traditional AI models often fail to capture the full complexity embedded within them. By relying mainly on the final scores from neuropsychological assessments, AI models reduce rich, multidimensional data into sparse representations that overlook key aspects of cognitive functioning. The diverse channels of data, such as response times, patterns of mistakes, and performance variations across modalities, hold significant potential that remains largely untapped. These oversights not only limit the accuracy of AD diagnosis but also restrict opportunities for early interventions tailored to individual cognitive profiles. To address this gap, our approach fully leverages the multimodal richness of neuropsychological test data, transforming

* Correspondence to: Chongqing University, No. 6 teaching building, district a, 174 shazheng street, Shapingba District, Chongqing, China.
E-mail address: qcui@cqu.edu.cn (Q. Cui).

https://doi.org/10.1016/j.cie.2024.110625
Received 12 June 2024; Received in revised form 20 September 2024; Accepted 1 October 2024
Available online 2 October 2024
0360-8352/© 2024 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

it into comprehensive feature sets that strengthen AI-based diagnostic models for AD.

Although utilizing neuroimaging data for diagnosing AD is a common practice (Kim, Jeong, Stiles, & Choi, 2022), its ability to reflect the full range of clinical features is limited. Beyond commonly observed features such as hippocampal atrophy, early-stage AD patients often exhibit declines in cognitive abilities related to language and social cognition (Harciarek & Cosentino, 2013), which neuroimaging alone cannot fully capture. This highlights the importance of neuropsychological tests that assess different cognitive domains in the AD diagnostic process. Existing research shows that memory-related functions, as assessed by tests like the Mini-Mental State Examination (MMSE) and Montreal Cognitive Assessment (MoCA) (Markwick, Zamboni, & de Jager, 2012), exhibit more pronounced differences between MCI and AD stages compared to non-memory cognitive domains (Loewenstein et al., 2006). Thus, simply combining tests from various cognitive domains without screening may introduce noise and reduce diagnostic accuracy. A more targeted approach, partitioning neuropsychological tests by specific cognitive domains, can provide a clearer, more precise evaluation of an individual's cognitive status, enhancing the diagnostic process.

Model fusion holds the potential to produce more stable and robust models to enhance diagnostic performance (Lei et al., 2019). Traditional model fusion techniques, such as voting (Mathew, Kumar, & Kumar, 2020), are influenced by the size of the training set, which significantly impacts the overall predictive performance of the ensemble model. Stacking, in particular, requires additional model design and often exhibits suboptimal performance due to ill-defined parameters (Prodromidis & Stolfo, 1999). These conventional methods exhibit limitations including instability, inefficient training, poor generalization, and a lack of interpretability (El-Sappagh, Alonso-Moral, Abuhmed, Ali, & Bugarín-Diz, 2023). Given the complexity of multimodal AD data, it is crucial for the model fusion framework to automatically allocate weights to different diagnostic models based on the characteristics of the data, rather than relying on subjective human selection (Eskandari, Aghaei, Milimonfared, & Nedaei, 2023). An innovative approach would involve learning weight parameters dynamically for each sub-model, thus improving both interpretability and performance. Additionally, this approach should assign different weights to data samples during fusion, similar to how clinical doctors focus more on challenging cases (Pan & Mishra, 2023). This method would iteratively address training errors thereby enhancing the final diagnostic accuracy. To date, no unified strategy exists for dynamically learning weight parameters and sample weights specifically for multimodal AD data.

Inspired by the previously discussed approaches, the embedding of multimodal clinical information contained in neuropsychological tests remains insufficient, and some of the cognitive domains included may not necessarily benefit AI models. Additionally, the design of model fusion strategies for integrating multimodal AD data is still relatively limited. Based on these considerations, this study seeks to explore the following four key questions:

(1) How can the multimodal clinical information embedded in neuropsychological tests be extracted?
(2) For multimodal AD data, how can the optimal degree of expansion be determined?
(3) How can diagnostic models be constructed from different cognitive domains within neuropsychological tests to eliminate the interference of redundant domains?
(4) How can a learnable model fusion strategy be designed to address the characteristics of multimodal AD data?

To answer these questions, this paper aims to present a multimodal AD diagnosis framework that expands neuropsychological test data using a polynomial kernel. It employs multiple Extreme Learning Machines (ELMs) to analyze various cognitive domains and uses a boosting algorithm to integrate model results and adjust weights for multimodal AD data. The contributions of this article are summarized as follows:

- We have developed a framework for a multimodal deep learning diagnostic model. This framework includes dimension expansion of neuropsychological tests and cognitive-based ELMs for cognitive domain analysis and selection. The framework dynamically adjusts model weights based on the errors of various models and the characteristics of multimodal AD data, ultimately boosting the overall diagnostic accuracy of the framework.
- We proposed a polynomial-based dimension expansion method to restore the rich information, including videos, speech, images, and text, which are embedded in neuropsychological tests, providing more comprehensive information for the AD diagnosis.
- We presented a series of cognitive-based ELMs to overcome redundancies and noise interference among cognitive domains within individual neuropsychological tests. These models can fully exploit the impact of cognitive domains on the AD and MCI diagnoses. We theoretically demonstrate that our proposed methods have better generalization capabilities compared to neuropsychological tests containing multiple cognitive domains.
- For base diagnostic models based on different dimensions and cognitive domains of multimodal AD data, we have designed a multimodal AD data-driven boosting ensemble method incorporating a modality factor. We theoretically prove that our proposed boosting method outperforms other ensemble methods.

The rest of this article is framed as follows: The literature review of the multimodal diagnostic algorithms based on neuropsychological tests is provided in Section 2. The problems description and assumptions within Section 3. The proposed solutions is elaborated in Section 4, followed by the results and discussions in Section 5 and the conclusions in Section 6.

2. Literature review

Following the analysis of multimodal AD data and neuropsychological tests in the previous section, this section briefly reviews the latest multimodal deep learning models based on neuropsychological tests, cognitive domain analysis, and ensemble model for AD.

2.1. Neuropsychological test feature processing

The majority of neuropsychological test total scores and scores for individual items are derived from evaluations involving speech, video, images, or text. Therefore, behind each score lies a wealth of clinical information. In recent years, multimodal AD diagnostic algorithms based on deep learning have demonstrated promising efficacy (Kim, Duong, & Gahm, 2024). Goel et al. (2023) proposed an RVFL classification model combined with ResNet-50 for feature extraction from multimodal AD data, however, its consideration of dimensions is limited. Ying, Yang, and Zhou (2023) introduced the use of Wav2Vec2.0 and BERT pre-trained models to construct speech and semantic features for AD patients, achieving good results on the ADReSSo task, yet this approach does not effectively utilize assessment results obtained from neuropsychological tests and does not incorporate neuroimaging information into fusion. Zhang, He, et al. (2023) proposed the MCAD framework for multimodal AD diagnosis, but the addition of the CSF encoder increases the computational burden of this framework, although it can transform CSF features from sparse to dense. Song, Mao, and Qian (2021) designed AMGNN, which combines the fitting ability of GNN to learn multimodal AD data. However, this method directly concatenates standardized scores from neuropsychological tests with neural imaging data, overlooking the numerous influence factors underlying neuropsychological test results. These issues highlight the growing demands on multimodal deep learning and underscore the importance of designing models that cover more modalities of AD data while densifying representations of non-euroimaging dimensions.

2.2. Cognitive domains for AD diagnosis

The integration of multiple cognitive domains in neuropsychological test assessments plays a crucial role in the diagnosis of AD. Gupta and Kahali (2020) proposed the use of random forest for accurately determining the status of cognitive domains such as semantic memory, perceptual orientation, and executive functioning, however, this method does not incorporate neural imaging data into the diagnostic model. Tosi et al. (2020)utilized network analysis to analyze relationships among various neuropsychological tests in AD patients and their associations with age and years of education. Although network analysis revealed the necessity of reorganizing the cognitive system, no diagnostic models based on reorganized neuropsychological tests have been developed. Bhasin, Ohri, Kumar, Sharma, and Thakur (2023) achieved promising results by selecting 13 memory-related features from 12 neuropsychological tests to build a machine learning model, primarily focusing on MCI patients, with limited cognitive domain analysis for AD patients. Therefore, these methods have not considered the redundancy, exclusivity, and overlap among cognitive domains in the tests. To the best of our knowledge, there is no research that has split individual neuropsychological test assessments based on cognitive domains to evaluate AD and MCI.

2.3. Multimodal AD diagnosis model fusion

Different diagnostic models have the capability to autonomously learn their diagnostic weights, and assigning varying weights to samples during training is a crucial ability of model fusion. Hakan (2024) proposed a soft voting method based on handwriting that automatically updates weights for incorrectly predicted samples to predict AD. However, their method lacks precision in ensemble parameter learning and does not consider a broader spectrum of AD data dimensions. While Krstev, Pavikjevikj, Toshevska, and Gievska (2022) proposed a predictive model for AD that combines acoustic, language, and image classification models using soft voting, and Shukla, Tiwari, and Tiwari (2024) utilized model fusion to investigate subcortical structures for AD detection, and Paplomatas, Krokidis, Vlamos, and Vrahatis (2023) introduced an early fusion approach for feature selection in Single-cell RNA sequencing (scRNA-seq) data focusing on AD-related analysis, these methods did not simultaneously consider the uniqueness of samples and the ability to autonomously learn ensemble parameters. Therefore, there is a need to design an automatic learning ensemble parameter model fusion strategy specifically for multimodal AD data that can mimic the differential diagnostic weight assignment by physicians to different samples. Additionally, in the boosting process, model error should not be the only factor determining model weights, the information content and variance of the input data also play a role (Rayana, Zhong, & Akoglu, 2016).

2.4. Research gaps

Although previous studies have achieved notable results, certain gaps remain. As shown in Table 1, with the advancement of deep learning, the dimensions involved in multimodal AD diagnosis have expanded significantly. In addition to the commonly used MRI/PET, other dimensions now include genetic data and neuropsychological scales. However, these studies have not fully explored the multimodal information within neuropsychological scales, nor have they further extended the features derived from them. Furthermore, there has been limited research focused on diagnostics based on cognitive domains included in neuropsychological scales. Most existing work still relies on vector concatenation for model fusion, and the fusion strategies designed for multimodal AD data remain relatively limited.

3. Problem description and assumptions

Building upon the gaps gained from the literature review, we now turn to a detailed examination of the multimodal AD diagnosis problems. This section outlines the specific problems we aim to address and the assumptions that underpin our approach.

3.1. The sparse neuropsychological features

Traditional neuropsychological assessments frequently fail to capture the full complexity of cognitive functioning. These assessments often yield sparse feature representations that do not adequately reflect the rich and multimodal information, such as video, speech, images, and text, which is crucial for accurate diagnosis. This limitation undermines the clinical utility of these assessments, potentially leading to misdiagnoses or delays in intervention. To overcome this limitation, we propose the use of a polynomial kernel function to achieve dimensional expansion.

For a specific multimodal AD dataset, the non-neuroimaging data is denoted as X,

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,k} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,k} \end{bmatrix}, \tag{1}$$

where n is the dataset size, and k denotes the total dimension of non-neuroimaging data. Non-neuroimaging consists of question items from different modalities. We aim to reconstruct a dense feature matrix using mathematical kernel, thereby restoring the multidimensional clinical information contained in these neuropsychological tests. Then, the transferred dataset dimension becomes: m (m > k), and

$$m = \phi(d, k), \tag{2}$$

the ϕ is not the linear function, the new dense feature becomes

$$\tilde{X} = \begin{bmatrix} \tilde{x}_{1,1} & \tilde{x}_{1,2} & \cdots & \tilde{x}_{1,m} \\ \tilde{x}_{2,1} & \tilde{x}_{2,2} & \cdots & \tilde{x}_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{x}_{n,1} & \tilde{x}_{n,2} & \cdots & \tilde{x}_{n,m} \end{bmatrix}. \tag{3}$$

The challenge can be transformed into solving the expanded dataset $\tilde{X}$.

3.2. The noise of cognitive domain

In a neuropsychological test, it is often possible to partition it into different subsets according to different cognitive domains. For example, the MMSE can be divided into subsets related to orientation, computation, attention, and others. Therefore, consideration can be given to segmenting the scale into different cognitive domains to generate distinct sub-datasets, and subsequently constructing diagnostic models based on these subsets. On the other hand, complete scale data encompass numerous item questions, some of which may contain noise, leading to diagnostic redundancies and increased susceptibility to interference. To overcome the issues, we propose to partition the data into different subsets, it can to some extent assist in identifying optimal classification data features from the scale, eliminating noisy features, and thereby facilitating the construction of more precise diagnostic models.

After partitioning the scale according to cognitive domains, constructing diagnostic models for these domains remains a challenge. The ELM presents a promising choice for this task. Firstly, after partitioning the scale, a considerable number of sub-datasets will be obtained. Due to the absence of gradient computation in ELMs, they can significantly enhance computational efficiency compared to other machine learning models. Moreover, ELMs are less prone to overfitting, which is attributed to their streamlined set of hyperparameters and reduced tendency to memorize noise present in the dataset. Additionally, ELMs

Table 1
Comparison of previous AD diagnosis works.

Algorithms	Dataset description	Modal	Neuropsychological scores	Cognitive domains	Fusion strategy
Gupta and Kahali (2020)	56AD+428MCI	Scales, Demographics	No extensions	Splitting	–
Tosi et al. (2020)	191AD+165CN	Scales, Demographics	No extensions	Splitting	–
Song et al. (2021)	337AD+856MCI+413NC	MRI,Scales, Demographics	No extensions	No splitting	Vector Concentrate
Zhu, Lu, Wang, Gorriz, and Zhang (2022)	97AD+233MCI+119NC	MRI	–	–	Average
Krstev et al. (2022)	257AD+52CN	Scales, Speech, Text	No extensions	No splitting	Voting
Song et al. (2022)	78AD+145NC	MRI,Demographics	–	–	Vector Concentrate
Li et al. (2022)	169AD+165MCI+168NC	MRI, Scales, Demographics	No extensions	No splitting	Vector Concentrate
Lin et al. (2023)	187AD+382MCI+229NC	MRI	–	–	Vector Concentrate
Paplomatas et al. (2023)	–	scRNA-seq	–	–	–
Goel et al. (2023)	210AD+210MCI+210CN	MRI/PET	–	–	Fusion rule
Ying et al. (2023)	343AD+378MCI	MRI/PET, Speech, Text	–	–	Vector Concentrate
Bhasin et al. (2023)	–	Scales	No extensions	Splitting	–
Rahim, El-Sappagh, et al. (2023)	182AD+100MCI+282NC	MRI, Scales, Demographics	No extensions	No splitting	Vector Concentrate
Zhang, He, et al. (2023)	129AD+125MCI+110NC	MRI/PET, Scales, CSF, Demographics	No extensions	No splitting	Vector Concentrate
Rahim, Abuhmed, Mirjalili, El-Sappagh, and Muhammad (2023)	182AD+100MCI+282NC	MRI/PET, Scales, CSF, Demographics	No extensions	No splitting	Vector Concentrate
Adarsh, Gangadharan, Fiore, and Zanetti (2024)	–	MRI,Scales, Demographics	No extensions	No splitting	–
Zhang, Cui, Lü, and Li (2024)	337AD+856MCI+413NC	MRI/PET, Scales, CSF, Demographics	No extensions	No splitting	Vector Concentrate
Kim et al. (2024)	84AD+208MCI+101CN	MRI/PET	–	–	Vector Concentrate
Hakan (2024)	89AD+85CN	Scales	No extensions	Splitting	Voting
Shukla et al. (2024)	200AD+200MCI+200CN	MRI/PET	–	–	Vector Concentrate
Proposed	54AD+109MCI+102NC	MRI/PET, Scales, CSF, Demographics	Extensions	Splitting	Enhanced boosting

showcase resilience to noise and outliers, a result of their stochastic initialization of hidden layer weights. This feature sets ELMs apart from traditional algorithms like Support Vector Machines (SVM) (Jakkula, 2006), Logistic Regression (LR) (LaValley, 2008), and Decision Trees (DT) (De Ville, 2013), which typically display heightened sensitivity to outliers and noisy features.

3.3. The influence of AD modalities

Based on the preceding analysis, partitioning neuropsychological tests by cognitive domains produces a series of base diagnostic models. In AD diagnosis, the weights assigned to results from different physicians often vary, with models of lower diagnostic accuracy typically receiving lower weights in the boosting process. However, due to the diverse contributions of various AD features, previous ensemble methods have not effectively accounted for the unique contributions of different cognitive domains and modalities. This oversight results in suboptimal diagnostic performance. Therefore, a more integrated approach that combines insights from multiple models and accounts for the contributions of different modalities is essential for improving the accuracy and reliability of AD classification.

3.4. Problem assumptions

In this subsection, we present the key problems and underlying assumptions that serve as the foundation of our study. To simplify the mathematical modeling process and experiments, several assumptions are made and listed as follows:

(1) All notations used in the following mathematical formulas are accurately explained and remain valid in the text following each formula.
(2) The finite set S contains a sufficient range of candidate degrees, ensuring that the optimal degree d should lie within this set.
(3) The proposed ELMs are single-layered with Gaussian activation functions.

(4) The outputs of the s ELMs obtained from $[D_1, D_2, \ldots, D_s]$ are mutually independent.
(5) The binary classification task of distinguishing between AD and non-AD can be extended to a three-class classification task, encompassing AD, MCI, and Cognitive Normal (CN).
(6) Use 1 to represent the positive class and −1 to represent the negative class.
(7) Our experimental dataset is sufficiently large to support the significance testing using the T-test.

4. Solution approach

To address the problems identified in prior section, we now turn to the details of our proposed solution approaches. The entire framework comprises four crucial components. The first is preprocessing module. The remaining three components include a broadening polynomial function for restoring multimodal information behind neuropsychological tests, a cognitive-based ELM focused on individual cognitive domains and a multimodal AD data-driven boosting ensemble of base diagnostic models combined with multimodal AD data.

4.1. Data preprocessing

During the acquisition of data from the Alzheimer's Disease Neuroimaging Initiative (ADNI), we selected patient data across several dimensions, including MRI, MMSE, CDR, MODHACH, NPIQ, ADAS, APOE, family history, demographics, and medical history. In the data preprocessing phase, we implemented additional steps specifically for MRI data. For embedding MRI data, we employed a fine-tuned Vision Transformer (ViT) model (Liu et al., 2021), known for its ability to capture long-range dependencies and global context in images, offering superior performance on large datasets compared to traditional convolutional neural networks. The data preprocessing was carried out in the following six steps:

Step 1: Slice the three-dimensional MRI data along different axes, obtaining slice data from three dimensions.

Step 2: For each set of slices, calculate the image entropy and rank them, with higher entropy indicating greater information content, thus being more helpful for observing pathological regions such as the hippocampus.

Step 3: Select the top 20 two-dimensional images from the ranked slices of each dimension, resulting in a total of 60 images.

Step 4: Load the ViT model weights and further fine-tune the model using the data obtained in Step 3 until convergence.

Step 5: Use the fine-tuned ViT model from Step 4 as the embedding model for MRI two-dimensional images.

Step 6: After embedding the 60 images for each sample, calculate the mean of the embeddings, which serves as the final MRI embedding.

4.2. Broadening polynomial

The use of a polynomial kernel allows for the transformation of input data into higher-dimensional space, enabling the model to capture more complex clinical patterns and relationships. Set the polynomial kernel is:

$$p_k = (w_i \cdot x_i \cdot w_j \cdot x_j + c)^d, \tag{4}$$

where i and j < k, p_k is the polynomial kernel, determining the degree d of the polynomial kernel can facilitate the transformation process. When lowdimensional data is mapped to a highdimensional space, if the data is independent in the highdimensional space, it will be beneficial for subsequent classification. To measure whether the independence of the transformed data from the original data has increased, we calculate the entropy of the entire dataset before and after transformation to quantify the change in its independence. The entire solution process comprises the following six steps:

Step 1: We calculate the entropy of the original data matrix and the entropy of the expanded data, denoted by H and $\tilde{H}$, respectively.

Step 2: We can derive the objective function:

$$O = \tilde{H} - H, \tag{5}$$

where O represents the object of the function, although a larger d leads to a higher dimensionality of the transformed data, it often leads to overfitting. Therefore, we need to find an optimal degree d that can simultaneously achieve the goal of expanding data dimensions and suppressing overfitting caused by excessive d.

Step 3: Accordingly, we introduce L1 (Schmidt, Fung, & Rosales, 2009) and L2 (Cortes, Mohri, & Rostamizadeh, 2012) regularization terms based on degree d, and subsequently construct a new objective function:

$$min\ (L) = \lambda_1 \cdot |d| + \lambda_2 \cdot |d|^2 - O, \tag{6}$$

Step 4: Calculate the gradient of d,

$$\frac{\partial L}{\partial d} = \lambda_1 + 2\lambda_2 \cdot d - \frac{\partial O}{\partial d} = 0, \tag{7}$$

which can find the best d. However, Object is a summation function and the upper bound of the summation terms is a function ϕ for d, it is therefore not possible to directly obtain the derivative of Object to d.

Step 5: Relax the constraints and formulate a new objective function,

$$L : S \to \mathbb{R}, \tag{8}$$

the S is a finite set containing b values, each of which has a minimum value greater than 1.

Step 6: Based on the above analysis, we can derive the expression for the optimal degree d as follows:

$$d^* = \{d_i \in S : L(d_i) = min_{d \in S} L(d)\}, \tag{9}$$

d^* represents the value of d when the objective function L achieves its minimum value, the optimal degree d of the transfer function in Eq. (4). Thus far, we have expanded the feature dimensions of neuropsychological tests using polynomial extension, while also constraining them to avoid overfitting caused by excessive dimensionality. The entire extension process can be referenced in Appendix A.1.

4.3. Cognitive-based ELM

In the ELM model, let there be U nodes in total, then the output y is given by:

$$y = \sum_{u=1}^{U} \boldsymbol{\beta}_y \mathbf{G}_y(x), \tag{10}$$

here, y represents the predicted value of the ELM, β denotes the parameter matrix of the ELM, and **G** is the output matrix of each node in the ELM (Huang, Zhu, & Siew, 2006). Let the complete scale data be denoted as **D**, after segmenting it according to the cognitive domain, s distinct subsets of data can be obtained, denoted as $[D_1, D_2, \ldots, D_s]$. For the sake of generality and to simplify the analysis, we assume that the ELMs within the model pool are single-layered, with Gaussian activation functions, the jth activation function at each basic ELM can be expressed as the product of all the corresponding Gaussian function,

$$g_o^j(x_o) = e^{\frac{-(x_o - a_{jo})^2}{b_j}}, \tag{11}$$

with the constraint $g_o^j(x_o) \subseteq (0,1)$ for the o-th original question items of the scale, a_{jo} and b_j are random values which are taken from same distributions, where o ∈ [1,2, …,r], then we can express the pth node for the ith basic ELM on sample x_q (q = 1,2, …,n) as

$$\boldsymbol{g}_{iq}^{p} = \prod_{v=1}^{r} g_{pv}^{i}(x_v^q), \tag{12}$$

based on Eq. (12), the **G** in Eq. (10) becomes

$$\mathbf{G}_i = (\prod_{v=1}^{r} g_{1v}^i(x_v^q), \prod_{v=1}^{r} g_{2v}^i(x_v^q), \ldots, \prod_{v=1}^{r} g_{uv}^i(x_v^q)). \tag{13}$$

Theorem 1. *The proposed strategy with ELMs for neuropsychological test classification, which relies on a subset of items extracted from the scale as Fig. 1, exhibits superior generalization capabilities compared to a single neuropsychological test classification method based on the entirety of the scale's items.*

Proof. See Appendix B.1.

4.4. Multimodal AD data-driven boosting

Considering the different influence of AD modalities, we design an AD modality factor z (z> 0) derived from the entropy and variance of the multimodal AD data input by different models. In our proposed boosting method, the AD modality factor enables weight adjustments by considering both model accuracy and the characteristics of the input data, thereby automatically learning an AD modality factor to adjust the weights. Additionally, the difficulty of diagnosis is also influenced by individual patient data. Compared to the Adaboost algorithm, the weight of its weak diagnosis classifier is denoted as

$$\alpha_0^t = \frac{1}{2} log \frac{1-\epsilon_t}{\epsilon_t}. \tag{14}$$

The ensemble diagnosis model of Adaboost is denoted as

$$H_0(\mathbf{X}) = \sum_{t=1}^{T} \alpha_0^t h_0(\mathbf{x}). \tag{15}$$

We designed a new weight update formula,

$$\alpha_1^t = \frac{1}{2} log \frac{1-\epsilon_t}{\epsilon_t} + \mathbf{z_t} e^{1-\epsilon_t}, \tag{16}$$

let ϵ_t represent the error of the t-th diagnostic model and z_t represent the AD modality factor of the t-th model. For any specific input of the model **x**, the AD modality factor is

$$z = \omega \cdot H(\mathbf{x}) + \tau \cdot \mathbb{D}(\mathbf{x}), \tag{17}$$

where ω and τ are learning parameters, H, $\mathbb{D}$ represent the entropy and variance of the input **x** respectively. Then, the proposed boosting model can be represented as

$$H_1(\mathbf{X}) = \sum_{t=1}^{T} \alpha_1^t h_1(\mathbf{x}). \tag{18}$$

The entire calculation process refers to the algorithm of Appendix C.1.

Theorem 2. *The proposed boosting method demonstrates superior true positive (TP) and lessened true negative (TN) rates respectively compared to Adaboost, indicating that the multimodal AD data-driven boosting possesses better generalization capability than Adaboost.*

Proof. See Appendix B.2.

The above analysis demonstrates that the proposed boosting algorithm can dynamically learn the weight parameters of various base diagnostic models and assign different weights to learning different samples. The entire integration process can be referenced in Fig. 2.

5. Results and discussions

After discussing the proposed solutions, this section begins by presenting our experimental data and settings. Next, polynomial functions of various degrees are compared with those derived from our proposed method. The results of constructing a single ELM using all non-neuroimaging data are then contrasted with the outcomes from our approach. Finally, the effectiveness of our ensemble method is evaluated against other commonly used ensemble techniques. To further demonstrate its robustness across different neuroimaging data perspectives, additional comparisons involving varying brain region sizes are provided.

5.1. Datasets

The dataset used in this study comes from the ADNI project, which is a subset of the ADNI database and contains multiple time point records and multimodal features for each sample. The selected samples include demographic information such as age, gender, and year of education. Additionally, neuropsychological test scores and MRI neuroimaging features are also included in the dataset. The dataset comprises records of 306 patients at various diagnostic time points. During data processing, the diagnostic data for the same patient at different time points were treated as separate samples. All samples must have values across all modalities. After screening, we included a total of 265 valid samples, with 54 AD, 109 MCI, and 102 CN samples.

In real clinical settings, the information provided by different angles of MRI slices varies significantly. To validate that our proposed methods can enhance diagnostic performance even in neuroimaging with insufficient perspectives in some brain regions, we used neuroimaging entropy as a reference and selected four groups of neuroimaging with different entropy values for experimentation as discussed in Section 4.1. In these four MRI datasets, entropy represents the clarity and spatial extent of the hippocampus depicted in the MRI images. Consultation with medical experts revealed that MRI images ranked below the top 90 based on entropy sorting could not provide meaningful diagnostic information. Consequently, we selected the top 90 MRI images for classification purposes, as shown in Fig. 3. A larger red area in the image indicates a better diagnostic perspective for the physician.

5.2. Experiment setup

During the process of constructing a pre-trained model for diagnosing neuroimaging, we selected the "google/vit-base-patch16-224-in21k" base model. The fine-tuning process was conducted using a GPU, completing 20 epochs of training. Subsequent neuroimaging feature extraction and ELM training were carried out on hardware consisting of an Intel i7 processor and 32 GB of RAM. All stages utilized Python 3.10.

To eliminate potential random factors in the experiments and minimize the impact of training and testing set divisions, we employed 5-fold cross-validation during the experimental process. Specifically, the entire dataset was divided into five subsets, and the experiment was repeated five times, with each subset serving as the test set in turn. The final experimental results were obtained by averaging the outcomes of these five iterations.

The evaluation encompasses six metrics. In the multimodal AD feature awareness stage, we employ the metrics of Accuracy, Precision, Recall (or Sensitivity), F1 Score, Area under Curve (AUC), and Matthews Correlation Coefficient (MCC). Accuracy measures the overall correctness of classifications, precision indicates the proportion of true positives among predicted positives, minimizing false positives, recall reflects the model's ability to identify all true positives, crucial for early detection, F1 score provides a balanced measure of precision and recall, and MCC offers a comprehensive assessment of model performance by considering true and false positives and negatives, especially useful for imbalanced datasets. These metrics collectively ensure the reliability and clinical relevance of the diagnostic model. We also calculate the T-test p-values related to accuracy-related experiments, during the calculation of these p-values, we computed the P-values between the proposed method and other control groups' test results, demonstrating significant differences in results between the proposed

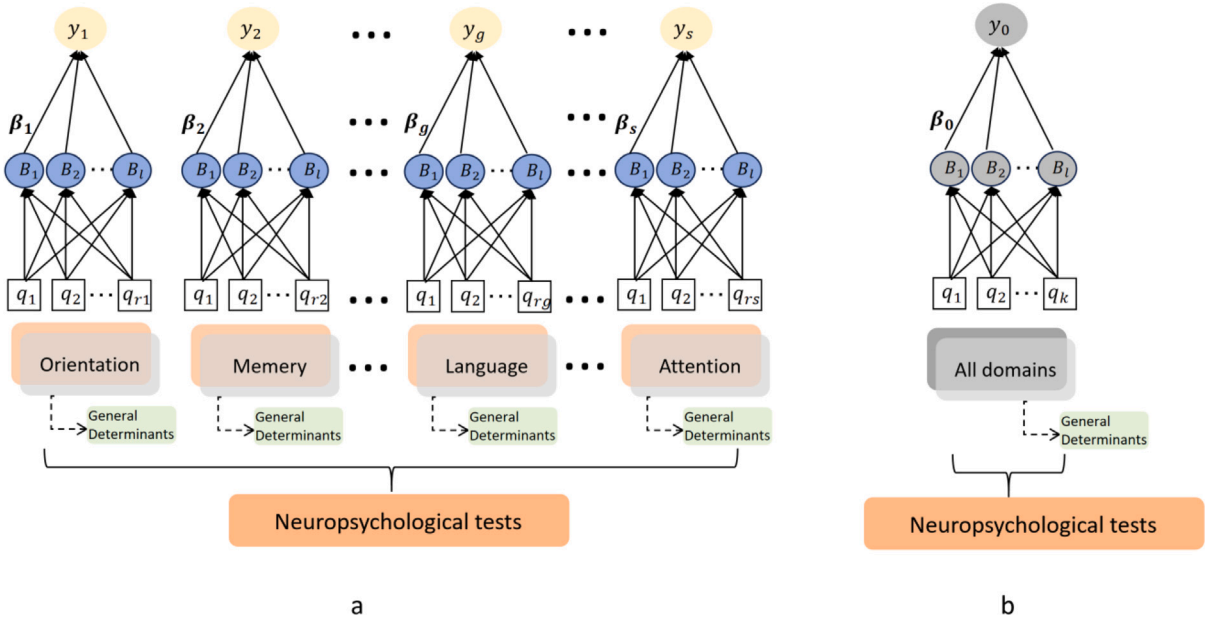

Fig. 1. The diagram of the cognitive-based ELM. "General determinants" refers to other dimensions of AD data, such as age and gender. Subgraph **a** as the proposed method, which involves constructing ELMs based on cognitive domains by segmenting various neuropsychological tests. In contrast, subgraph **b** represents not segmenting neuropsychological tests and directly constructing ELMs.

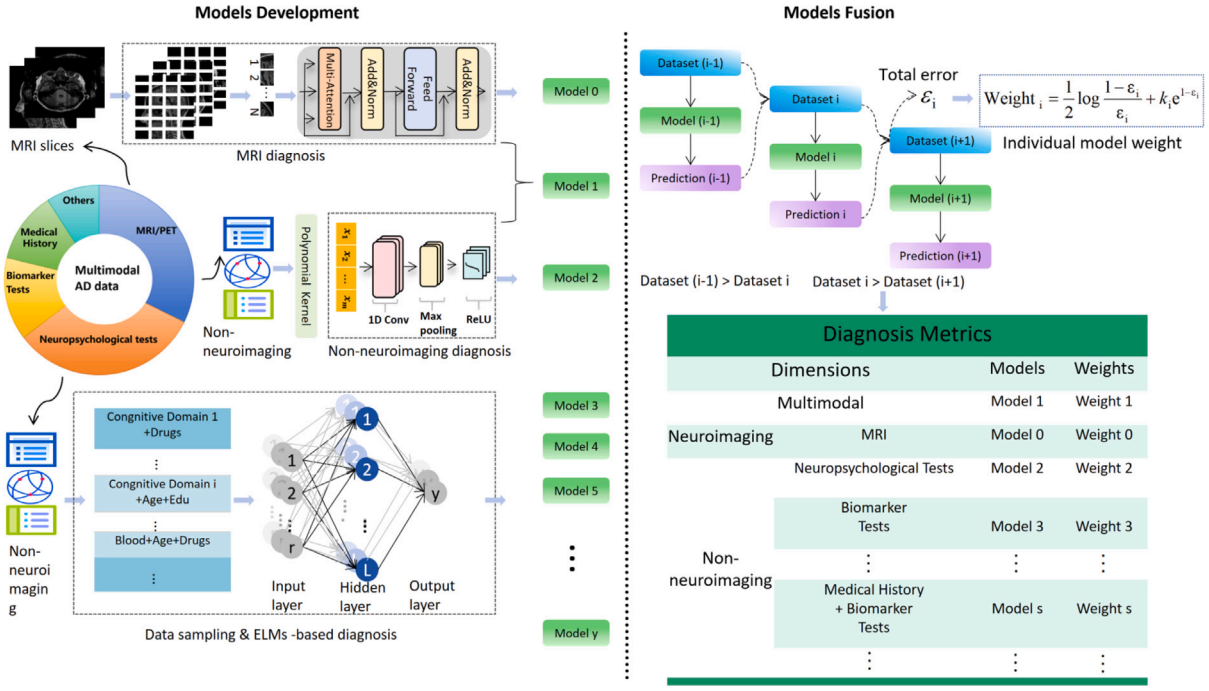

Fig. 2. Overall structure of the framework.

methods and control groups. If a statistically significant difference in P-values is observed between different experimental groups, it can be concluded that the accuracy achieved is unlikely to be due to random chance. Besides, we conduct separate assessments for Precision, Recall, F1 Score, and AUC values for each of the three classes—CN, MCI, and AD. The specific definitions are as follows:

$$\text{Accuracy} = \frac{TP + TN}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \tag{19}$$

$$\text{Precision} = \frac{TP}{TP + FP}. \tag{20}$$

$$\text{Recall} = \frac{TP}{TP + FN}. \tag{21}$$

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{22}$$

$$\text{AUC} = \int ROC(f)df, \tag{23}$$

in this context, ROC(f) denotes the Receiver Operating Characteristic (ROC) curve based on the threshold f.

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}, \tag{24}$$

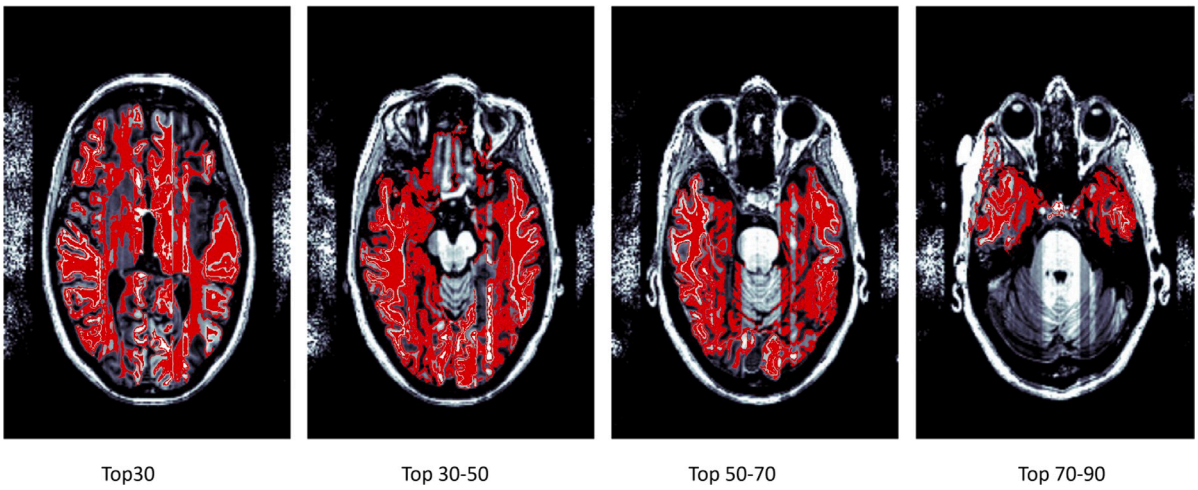

Fig. 3. 4 groups of MRI images with different entropy levels. The terms "Top 30" "Top 30-50" "Top 50-70" and "Top 70-90" represent four groups within a single patient's MRI data, each denoting the values of entropy within specific ranges (i.e., the top 30, top 30-50, top 50-70, and top 70-90 percentiles) across individual slices. A higher entropy value indicates a greater amount of information that the neuroimaging can provide. The regions marked in red indicate the brain areas of interest for the subjects.

Table 2
The overall and individual group performance of polynomial expansions of different degrees in ELM.

Polynomialdegree	Class	F1 (%)	Precision (%)	Recall (%)	CN vs. MCI vs. AD (%)		
					Accuracy	F1	MCC
d = 1[1]	CN	93.87	88.46	100	86.79	86.27	78.53
	MCI	83.72	85.71	81.82			
	AD	71.42	83.33	62.50			
d = 2[2]	CN	93.87	88.46	100	86.79	85.85	78.50
	MCI	84.44	82.61	86.36			
	AD	66.67	100	50			
Proposed[3] d = 3	CN	100	100	100	90.56	88.9	85.48
	MCI	89.79	81.48	100			
	AD	54.54	100	37.50			
d = 4[4]	CN	93.87	88.46	100	83.02	80.34	72.36
	MCI	80.85	76	86.36			
	AD	40	100	25			
d = 5[5]	CN	92	85.19	100	81.13	79.54	69.03
	MCI	77.27	77.27	77.27			
	AD	50	75	37.50			

Note. $P_{13} = 4.16 \times 10^{-3}$, $P_{23} = 4.07 \times 10^{-2}$, $P_{34} = 7.29 \times 10^{-4}$, $P_{35} = 3.28 \times 10^{-5}$ as derived from a T-test.

where false positives are indicated by FP, and false negatives are expressed as FN. With the experimental setup established, next, we delve into the analysis and discussions of our numerical results.

5.3. Performance of broadening polynomial

To validate the precision of our proposed method in identifying the optimal degree within a polynomial, we manually compared the diagnostic performance using ELM when degrees were set at 1, 2, 3, 4, and 5. In our experiments, we observed that as the degree exceeded 5, the dimensional expansion of neuropsychological test features became excessively large, leading to computer memory issues such as "Out of memory" situations. Analyzing the results from Table 2, we noted that at degree 3 (consistent with our proposed method), the accuracy, F1 score, and MCC were all higher compared to other degrees in the CN vs. MCI vs. AD three-class classification. The proposed method demonstrated consistent performance across the three metrics: accuracy, F1 score, and MCC, at degree 3. Moreover, it was observed that the accuracy metric significantly differed among the groups, with all p-values obtained from the paired t-test being less than 0.05. Despite the p-values for d = 2 and d = 3 being very close to 0.05, there still existed a statistically significant difference between them. Therefore, the *P*-value from the T-test further confirms that the accuracy in our experiment is significantly different from that of the control groups and not due to random chance. Beyond degree 3, there was a decreasing trend in overall diagnostic performance, aligning with the objective function L computed in Eq. (6).

In summary, the proposed broadening polynomial achieved an accuracy of 90.56%, an F1 score of 88.9%, and an MCC of 85.48%, demonstrating strong performance across multiple metrics. In clinical practice, this high accuracy indicates the model's potential to reliably distinguish between AD, MCI, and CN patients, reducing the likelihood of misdiagnosis. The F1 score, balancing precision and recall (or sensitivity), suggests the model's effectiveness in both identifying AD patients and minimizing false positives, which is crucial for timely intervention. Furthermore, the MCC reflects consistent classification across all three categories, ensuring that the model performs well not only for AD but also for MCI and CN, supporting patient stratification and tailored treatment strategies.

Upon examining the diagnostic outcomes for CN, MCI, and AD groups, we found that although degree 3 yielded the best overall performance, the F1 score for the AD group was not the highest. Instead, it was highest for the CN and MCI groups, which is in line with previous research indicating that neuropsychological tests are primarily

Table 3
The proposed cognitive-based ELMs compare with one ELM for overall and individual group performance.

Methods	Class	F1 (%)	Precision (%)	Recall (%)	CN vs. MCI vs. AD (%)		
					Accuracy	F1	MCC
One ELM[6]	CN	97.78	100	95.65	94.33	94.22	91.06
	MCI	93.62	88	100			
	AD	85.71	100	75			
Proposed[7]	CN	100	100	100	98.11	98.07	96.98
	MCI	97.78	95.65	100			
	AD	93.33	87.5	100			

Note. $P_{67} = 3.62 \times 10^{-5}$ as derived from a T-test.

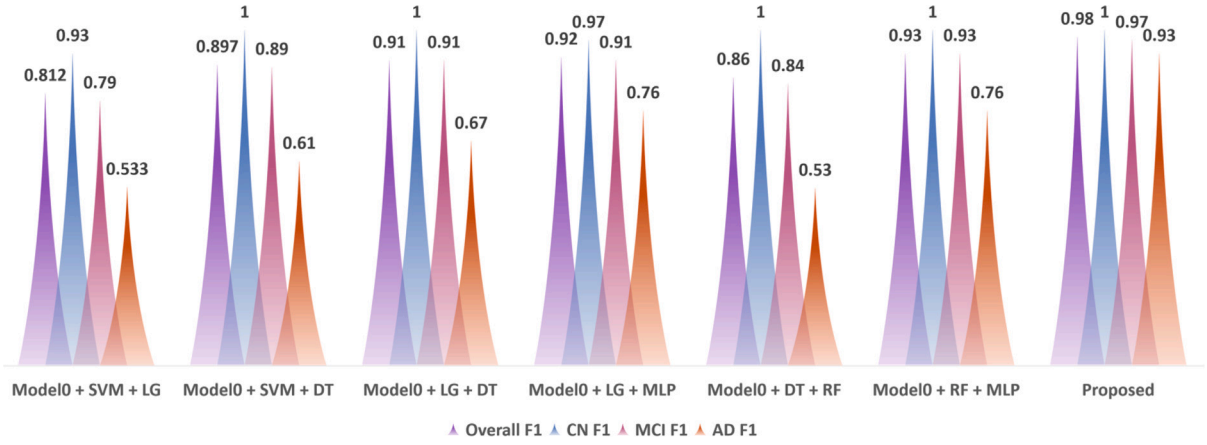

Fig. 4. Comparison of neuroimaging integration with machine learning algorithms: overall and individual F1 scores. Model 0 represents the diagnostic model constructed from neuroimaging dimensional data.

useful for assessing non-AD groups (e.g., MCI). Diagnosing AD requires incorporating additional dimensions such as neuroimaging alongside neuropsychological tests (Zhang et al., 2024). However, our proposed method still captures the multidimensional assessment processes inherent in neuropsychological tests, thereby improving performance for the CN and MCI groups.

5.4. Performance of the cognitive-based ELM

Our proposed cognitive-based ELMs integrate numerous neuropsychological tests by categorizing them based on cognitive domains, aiming to construct an integrated diagnostic model for multimodal AD data. To validate the effectiveness of this proposed method, we first compared it with a single ELM method, as shown in Table 3. Using our proposed cognitive-based ELMs, the accuracy, F1 score, and MCC metrics for the CN vs. MCI vs. AD classification task improved by 3.78%, 3.85%, and 5.92%, respectively. In other words, the proposed cognitive-based ELM model demonstrated consistent performance across the three metrics: accuracy, F1 score, and MCC, further validating its effectiveness. The significance of accuracy ($P < 0.05$) was determined using a paired t-test, indicating a significant difference between the results of our proposed cognitive-based ELM and the single ELM group. This further confirms that the accuracy obtained is valid and not due to random variation. Analyzing the results for individual CN, MCI, and AD groups, our proposed cognitive-based ELMs not only further improved CN and MCI classifications but also increased the F1 score for AD by 7.62%. These comparisons collectively demonstrate that when combined with neuroimaging, our proposed cognitive-based ELMs effectively enhance the diagnostic accuracy of multimodal AD.

In addition to applying all neuropsychological tests to a single ELM, we also compared our proposed cognitive-based ELMs with other machine learning models. We selected SVM, LR, DT, random forest (RF), MLP, and neuroimaging (referred to as Model 0) for ensemble modeling because these models represent a diverse range of commonly used algorithms with varying strengths in classification tasks (Shafiei, Tatar, Rayhani, Kairat, & Askarova, 2022). This selection allows for a comprehensive evaluation of our proposed model's performance across different methodologies and provides a robust comparison to assess its effectiveness. Then we computed the overall F1 score for CN vs. MCI vs. AD three-classification tasks and the individual F1 scores for each group, as shown in Fig. 4. The F1 score demonstrated consistent superiority across all four comparisons, the total F1 score achieved by the proposed method reached 98%, surpassing all other combinations. Moreover, our proposed approach is the only one that could elevate the F1 score for the AD group to over 90%. In contrast, methods such as SVM, LR, DT, and RF yielded AD group F1 scores below 60%.

These comparative experiments strongly indicate that our proposed cognitive-based ELM method effectively harnesses useful information from neuropsychological tests, outperforming common machine learning approaches in terms of diagnostic performance. Moreover, These high values indicate the model's exceptional ability to accurately classify AD, MCI, and CN patients, significantly reducing the risk of misdiagnosis. Clinically, such performance supports reliable early diagnosis and precise patient stratification, facilitating timely and effective intervention strategies.

5.5. Performance of multimodal AD boosting

Multimodal AD data encompass various modalities of information. To address this, we designed the multimodal AD data-driven boosting method. To validate its effectiveness, we compared it with Voting (Leon, Floria, & Bădică, 2017), Bagging (Galar, Fernandez, Barrenechea, Bustince, & Herrera, 2011), Stacking (Kwon, Park, & Lee, 2019), and Adaboost methods (Hambali, Saheed, Oladele, & Gbolagade, 2019), with Stacking using a random forest as the meta-model. These ensemble algorithms are commonly used in current research and have

Table 4
Comparison of performance between proposed boosting and ensemble methods: voting, bagging, stacking, and adaboost.

Methods	Class	F1 (%)	Precision (%)	Recall (%)	CN vs. MCI vs. AD (%)		
					Accuracy	F1	MCC
Voting[8]	CN	93.87	88.46	100	84.90	83.99	75.30
	MCI	81.81	81.81	81.81			
	AD	61.53	80	50			
Bagging[9]	CN	100	100	100	90.56	88.90	85.48
	MCI	89.79	81.48	100			
	AD	54.54	100	37.50			
Stacking[10]	CN	95.45	100	91.30	86.79	86.81	78.75
	MCI	85.10	80	90.90			
	AD	66.67	71.43	62.50			
Adaboost[11]	CN	100	100	100	96.22	96.03	94.02
	MCI	95.65	91.67	100			
	AD	85.71	100	75			
Proposed[12]	CN	97.87	95.83	100	98.11	98.07	96.97
	MCI	100	100	100			
	AD	93.33	100	87.50			

Note. $P_{8,12} = 1.92 \times 10^{-5}$, $P_{9,12} = 3.87 \times 10^{-4}$, $P_{10,12} = 7.25 \times 10^{-5}$, $P_{11,12} = 5.13 \times 10^{-3}$ as derived from a T-test.

demonstrated strong performance in various AI model integration tasks. From Table 4, it is evident that our proposed boosting method achieves the highest accuracy of 98.11% in the three-classification task of CN vs. MCI vs. AD. The *P*-value from the T-test, which is less than 0.05, further validates the accuracy of this result. Additionally, the consistent findings across accuracy, F1 score, and MCC metrics indicate that our proposed method outperforms other common ensemble methods. In contrast, the accuracy of Voting and Stacking ensemble methods is below 90%. Furthermore, the F1 score and MCC of proposed boosting method are also higher than the other four ensemble methods. These high metrics underscore the model's exceptional performance in distinguishing between CN, MCI, and AD, thereby enhancing diagnostic precision and minimizing the risk of misdiagnosis. Clinically, such accuracy and consistency support effective early intervention and tailored treatment strategies, improving patient management and outcomes.

For the AD group, Voting, Bagging, and Stacking ensemble methods were unable to elevate the F1 score to 70%, whereas both boosting methods achieved 85.71% and 93.33% F1 scores, respectively. Compared to Adaboost, although the proposed boosting method slightly decreases the F1 score for the CN group, it still exceeds 95%. Moreover, it enhances the F1 scores for the MCI and AD groups by 4.35% and 7.62%, respectively. This indicates that updating the weights of each diagnostic model in multimodal AD data by increasing the weights of models with high diagnostic performance can improve the overall ensemble performance.

5.6. Sensitivity analysis

Based on the perspective of sensitivity analysis, our proposed framework should be effective across MRI images with varying spatial extents and clarity levels. From the results in Table 5, we observed varying degrees of improvement in each of the four groups under our designed ensemble strategy. In terms of accuracy, F1 score, and MCC metrics, the proposed ensemble method demonstrated consistent improvements across the four different experimental groups. This indicates that our proposed method is effective for MRI images of the hippocampal region, regardless of varying spatial extents. Particularly in the Top 70-90 group, using the proposed ensemble methods increased the accuracy for CN vs. MCI vs. AD by 16.98%, F1 score by 16.51%, and MCC by over 20%. However, in the Top 30-50 group, we also found a decrease of 7.62% in the F1 score for AD after using ensemble methods. Besides, in the groups Top 30-50, Top 50-70, and Top 70-90, the P-values of accuracy were all less than 0.05. Therefore, the differences observed among these three experimental groups are statistically significant, confirming the reliability of the accuracy results. In contrast, in the Top 30 group, the *P*-value exceeded 0.05. One plausible explanation for this observation may stem from the extensive clinical information provided by MRI within the Top 30 group, facilitating relatively straightforward diagnosis. In other words, using MRI data alone for diagnosing class AD is already highly effective. Incorporating additional dimensions of AD data may actually complicate the diagnosis of class AD. Consequently, in the Top 30 group, there appeared to be no significant difference between the performance of ensemble and non-ensemble methods. As analyzed earlier, adding neuropsychological tests is primarily beneficial for diagnosing CN and MCI groups, which suggests that ensemble methods may introduce some noise interference for the AD group in this experimental group.

By segmenting the MRI images based on clarity for sensitivity analysis, our proposed AI method demonstrates effectiveness across varying levels of image quality. Clinically, this suggests that our approach is robust and reliable for diagnosing AD regardless of the clarity of MRI images, ensuring consistent performance in diverse practical settings.

After obtaining the final ensemble results, We compiled the weight and other relevant information of all primary diagnostic base models in Table 6. In the analysis, the table highlights the top 36 diagnostic base models out of a total of 64. These analyses help explore the correlations and significance among different modalities of AD data, providing valuable insights for future efforts to streamline the AD diagnostic process. The main conclusions we have drawn are as follows:

- Model 0 represents the neuroimaging dimension, which is critical for AD diagnosis, as it alone contributes 25% to the final decision-making process. This highlights the dominant role of neuroimaging in accurately identifying AD.
- Model 1 combines neuroimaging with polynomial expansions of neuropsychological test features, contributing 10% to the overall model. This integration underscores the added value of combining imaging with cognitive data for improved diagnostic accuracy.
- Other models, particularly those relying on individual neuropsychological tests or cognitive domains, had lower weights, indicating that these modalities alone do not have a significant impact on the diagnosis. However, they can still provide supplementary information when integrated with neuroimaging.

These weight distributions suggest that while neuropsychological tests alone are insufficient for precise AD diagnosis, their combination with neuroimaging leads to a more comprehensive and accurate result. Additionally, the models with lower weights correspond to specific test items, implying that some cognitive domains may be redundant and could be eliminated to streamline the diagnostic process without sacrificing accuracy.

Table 5
Classification comparison within 4 distinct levels of neuroimaging in terms of information entropy.

Neuroimaging group		Class	F1 (%)	Precision (%)	Recall (%)	CN vs. MCI vs. AD (%)			
						Accuracy	F1	MCC	AUC
Top 30[13]	Without ensemble	CN	95.65	95.65	95.65	96.22	96.22	93.87	99.52
		MCI	100	100	100				
		AD	87.5	87.5	87.5				
	With ensemble	CN	100	100	100	98.11	98.07	96.98	–
		MCI	97.78	95.65	100				
		AD	93.33	87.5	100				
Top 30-50[14]	Without ensemble	CN	93.33	95.45	91.3	92.45	92.49	87.82	99.56
		MCI	91.3	87.5	95.45				
		AD	93.33	100	87.5				
	With ensemble	CN	100	100	100	96.22	96.03	94.01	–
		MCI	95.65	91.67	100				
		AD	85.71	100	75				
Top 50-70[15]	Without ensemble	CN	81.81	85.71	78.26	83.02	82.69	72.66	94.71
		MCI	85.71	77.78	95.45				
		AD	76.92	100	62.5				
	With ensemble	CN	86.95	86.95	86.95	86.79	86.76	78.42	–
		MCI	86.95	83.33	90.91				
		AD	85.71	100	75				
Top 70-90[16]	Without ensemble	CN	78.26	78.26	78.26	75.47	75.47	60.21	93.72
		MCI	77.27	77.27	77.27				
		AD	62.5	62.5	62.5				
	With ensemble	CN	97.87	95.83	100	92.45	91.98	87.85	–
		MCI	91.3	87.5	95.45				
		AD	76.92	100	62.5				

Note. $P_{13} = 6.29 \times 10^{-2}$, $P_{14} = 5.64 \times 10^{-4}$, $P_{15} = 1.98 \times 10^{-4}$, $P_{16} = 3.02 \times 10^{-5}$ as derived from a T-test. The results obtained after the ensemble process do not include predicted probability values. therefore, AUC values are not computed for any of the ensemble experimental groups.

Table 6
Comparison of primary diagnostic model weights, percentages, and their corresponding modal and domains.

Modals	Domains	Model index	Weights	Percentages
All modals	All	1	0.24	10%
Neuroimaging	–	0	0.61	25%
MMSE+Medical history	Orientation,Attention,Memory,Others	2,3,4,5,6	0.29	12%
MMSE+CBB	Orientation,Community affairs,Others	7	0.03	1%
CDR+CBB	Memory,Language,Others	8	0.02	1%
MODHACH+Biomarker	Behavioral,Daily activities,Others	9	0.07	2%
NEUROBA+Others	Psychosocial functioning,Behavioral symptoms,Others	10,11	0.2	8%
NPIQ	Mood and Emotion,Hallucinations and delusions,Others	12,13	0.18	7%
NPIQ+ADAS	Behavioral disturbances,Language,Construction,Others	14	0.08	3%
Random combination	–	15,16,...,35	0.64	27%

Note. The term "Random combination" denotes a random selection from all non-neuroimaging modalities, grouped based on cognitive domains. Each model selects 2-3 cognitive domains to construct the model. In the "Others" category under the "Modals" column, it indicates certain general determinations such as gender, while in the "Others" category under the "Domains" column, it represents specific cognitive domains.

5.7. Practical and theoretical implications

The findings of this study carry significant practical and theoretical implications for the field of multimodal AD diagnosis and cognitive assessment. Practically, the proposed multimodal framework enhances the accuracy and reliability of AD and MCI diagnoses, providing clinicians with a more robust tool for early detection and intervention. This advancement can lead to improved patient outcomes through timely and tailored therapeutic strategies. However, our validation has so far been limited to the complete ADNI dataset. In real clinical settings, data may not be as comprehensive, for instance, some patients may have missing scale data, resulting in sparse input data for the model. This sparsity could potentially reduce the framework's performance. Therefore, before application, the proposed framework requires further validation and iterative optimization with more extensive datasets. Theoretically, the research contributes to the understanding of cognitive domain interactions and their impact on diagnostic processes, offering a novel perspective on the integration of diverse data modalities. By demonstrating the efficacy of cognitive-based ELMs and a data-driven boosting ensemble approach, this study lays the groundwork for future research aimed at refining diagnostic models and exploring the complexities of cognitive functioning in neurodegenerative diseases.

6. Conclusions and future prospects

Building on prior research, conventional CAAD methods often struggle to capture the rich, multidimensional data inherent in neuropsychological assessments. This limitation results in a reliance on isolated scores that may not fully reflect the nuanced clinical context. Additionally, previous approaches have not adequately filtered and removed redundant cognitive domains from these assessments. When integrating models, it is important to consider not only model performance but also the structure of multimodal AD data. These challenges hinder early detection and intervention, highlighting the need for more advanced diagnostic techniques capable of integrating diverse data modalities.

This study not only fills existing gaps in the literature but also offers strong motivation for the use of advanced AI techniques in clinical practice. We propose the polynomial kernel and cognitive-based ELM to expand and filter the clinical information from neuropsychological assessments. Additionally, the designed boosting algorithm not only accounts for model differences but also incorporates the entropy and variance of multimodal AD data into the model's weight update process. Our proposed framework achieved a diagnostic accuracy of 98% on the ADNI dataset, with F1 scores of 93% and 97% for the AD and MCI groups, respectively, and an impressive 100% accuracy for the CN group. Moreover, sensitivity analysis demonstrates that the framework maintains its effectiveness across varying spatial extents and levels of clarity. Furthermore, we quantified the weights of sub-models corresponding to different modalities of AD data, providing new insights not only for advancing the application of CAAD but also for simplifying research on neuropsychological assessments.

While our proposed framework demonstrates high diagnostic accuracy on the ADNI dataset, it is essential to acknowledge certain limitations that may affect its broader applicability. One potential shortcoming is the generalizability of the model to other datasets, especially those with different population characteristics, imaging protocols, or cognitive assessment formats. The framework's performance may vary when applied to diverse clinical settings, and further validation on external datasets is necessary to confirm its robustness. Additionally, our model does not explicitly address the challenges posed by missing data, which can be a common issue in real-world clinical scenarios. Future research should explore strategies to handle incomplete data effectively, such as generative AI (GAI) or synthesis strategy based on GAI, to ensure the framework's applicability in a wider range of clinical environments.

List of abbreviations

AD	Alzheimer's Disease
ADNI	Alzheimer's Disease Neuroimaging Initiative
AI	Artificial Intelligence
MRI	Magnetic Resonance Imaging
MCI	Mild Cognitive Impairment
CN	Cognitive Normal
CDR	Clinical Dementia Rating
AUC	Area Under Curve
MCC	Matthews Correlation Coefficient
RF	Random Forest
MLP	Multilayer Perceptron
LR	Logistic Regression
SVM	Support Vector Machine
CBB	Cogstate Brief Battery
MMSE	Mini Mental State Examination
ADAS	Alzheimer's Disease Assessment Scale
MODHACH	Modified Hachinski Scale
NPIQ	The Neuropsychiatric Inventory Questionnaire

CRediT authorship contribution statement

Meiwei Zhang: Writing – original draft, Software, Methodology, Conceptualization. **Qiushi Cui:** Writing – review & editing, Investigation. **Yang Lü:** Resources. **Weihua Yu:** Resources. **Wenyuan Li:** Writing – review & editing, Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Original embedding pkl (sagittal plane, coronal plane, transverse profi le, etc.) (Original data)

Declaration of generative AI in scientific writing

The authors retain full responsibility and accountability for the entirety of this work, affirming the absence of generative AI in its conception and development.

Acknowledgments

This work was funded by the Science Innovation Programs Led by Academicians in Chongqing and the Science and Technology Research Program of Chongqing Education Commission, under Award Numbers 2022yszx-jsx0002cstb and KJQN201900109, respectively.

Appendix A. Appendix figures

A.1. Diagram of the broadening polynomial

See Fig. 5.

Appendix B. Appendix proofs

B.1. Proof of Theorem 1

Proof of model bias: Consider an arbitrary cognitive domain subset D_s generated from the complete scale **D**, where the feature dimension is denoted by r (r < k). Within subset D_s. If a basic ELM has U Gaussian hidden nodes, then the output of this ELM is

$$y_s = \sum_{j=1}^{U} \beta_j g_j(\boldsymbol{X_s}) = \sum_{j=1}^{U} \beta_j [g_1^j(x_1) \cdot g_2^j(x_2) \cdots g_r^j(x_r)], \tag{25}$$

where $\boldsymbol{\beta_j} = [\beta_1, \beta_2, \ldots, \beta_k]$. Similarly, if the cognitive domain of the scale is not divided, and ELM is directly applied to all items (**D**), then the output is

$$y_0 = \sum_{j=1}^{U} \beta_j g_j(\boldsymbol{X_0}) = \sum_{j=1}^{U} \beta_j [g_1^j(x_1) \cdot g_2^j(x_2) \cdots g_k^j(x_k)], \tag{26}$$

set

$$\mu_j = g_{r+1}^j(x_{r+1}) \cdot g_{r+2}^j(x_{r+2}) \cdots g_k^j(x_k), \tag{27}$$

according to the defined Gaussian function, $\mu_j \in (0, 1)$, then

$$y_0 = \sum_{j=1}^{U} \beta_j g_j(\boldsymbol{X_s}) \cdot \mu_j. \tag{28}$$

Let the model bias of an ELM model trained on a cognitive domain dataset D_s separated from the complete scale be denoted as $b_s = \bar{y} - y_s$, where $\bar{y}$ represents the corresponding actual output, and the model bias of an ELM trained on the complete scale dataset be denoted as $b_0 = \bar{y} - y_0$,

$$\begin{aligned} b_s &= \bar{y} - y_s = \bar{y} - y_0 + y_0 - y_s \\ &= b_0 + y_0 - y_s = b_0 + \sum_{j=1}^{U} \beta_j g_j(\boldsymbol{X_s}) \cdot \mu_j - \sum_{j=1}^{U} \beta_j g_j(\boldsymbol{X_s}) = b_0 \\ &\quad + \sum_{j=1}^{U} \beta_j g_j(\boldsymbol{X_s}) \cdot (\mu_j - \boldsymbol{I}), \end{aligned} \tag{29}$$

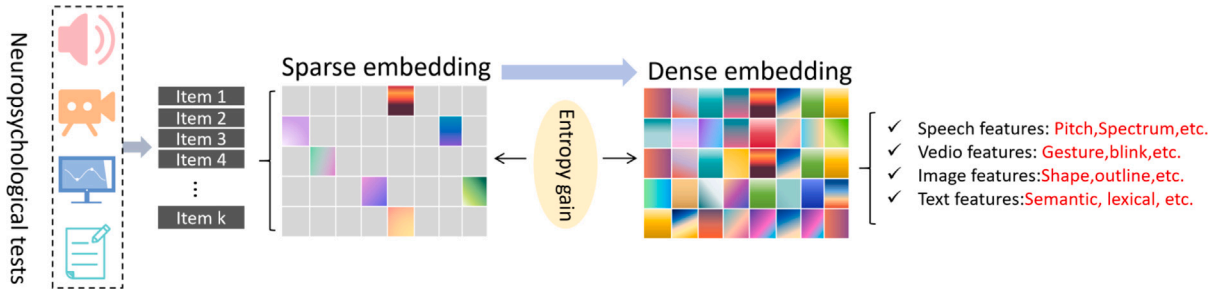

Fig. 5. The schematic diagram of the broadening polynomial.Although the original scores of the neuropsychological scale are derived from speech, video, images, and text, the resulting feature matrix is sparse. After applying our designed broadening polynomial, the feature matrix becomes dense, allowing it to represent the rich information underlying the problems of each modality.

since $\mu_j \in (0,1)$, so $(\mu_j - 1) < 0$ always hold true, so the last item in Eq. (29) from the monotonic linear function concerning β_j,

$$b_s < b_0. \tag{30}$$

which is always in existence, indicating that, unlike b_0, the proposed method can enhance the generalization capability of β_j by leveraging b_s.

Proof of model variance: Let the model variance of an ELM model trained on a set of cognitive domain datasets $[D_1, D_2, \ldots, D_s]$ generated from the complete scale be denoted as v, and the model variance of an ELM trained on the complete scale dataset **D** be denoted as v_0. Due to the properties of variance for a constant c,

$$\mathbb{D}(cX) = \mathbb{E}[(cX - \mathbb{E}[cX])^2] = c^2\mathbb{E}[(X - \mathbb{E}[X])^2] = c^2\mathbb{D}(X). \tag{31}$$

As there are no cross-item interactions within each cognitive domain dataset, it can be assumed that the outputs of the s ELMs obtained from $[D_1, D_2, \ldots, D_s]$ are mutually independent.

$$\mathbb{D}(y_1 + \cdots + y_s) = \mathbb{D}(y_1) + \cdots + \mathbb{D}(y_s). \tag{32}$$

Regardless of the subsequent ensemble strategy employed for integrating the outputs of the s ELMs, each ELM will have a weighting parameter less than 1. For the sake of simplicity in illustration, the averaging ensemble strategy is adopted here to demonstrate

$$v = \mathbb{D}(\frac{1}{s}\sum_{i=1}^{s} y_i) = \frac{1}{s^2}\mathbb{D}(\sum_{i=1}^{s} y_i) = \frac{v_0}{s}, \tag{33}$$

that is, given v is less than v_0, besides, compared to the classical Bootstrap sampling method, the proposed method can avoid the issue of Out Of Bag (Khan et al., 2021) because for any scale, the probability of any item feature being sampled is 100%.

B.2. Proof of Theorem 2

Within this section, we convert a multimodal AD dataset with three classes: AD, MCI, and CN into three binary classification problems. We assume that N_{ad} and N_{non-ad} are the number of AD and non-AD (MCI and CN) samples, respectively, where

$$N_{ad} + N_{non-ad} = N, \tag{34}$$

the parameters z and δ is determined by the training data

Definition 1. Set

$$\delta = \frac{N_{non-ad}}{N_{ad}}, \tag{35}$$

where $0 < \delta \leq 1$. For the t-th iteration, the sample weight is $D_t(i)$, i=1, ...,n. Let two new variables

$$\begin{cases} A_t = \sum_{i:y_i=1} D_t(i) \\ T_t = \sum_{i:y_i=1,h_t(x)=1} D_t(i), \end{cases} \tag{36}$$

then set

$$\varrho = \frac{T_t}{A_t}. \tag{37}$$

If we consider AD as positive, then TP indicates that the sample's label is actually positive and the predicted result H(x) is also positive. TN indicates that the sample's label is actually negative and the predicted result H(x) is also negative. A better diagnosis classifier will have higher TP and lower TN values respectively. Proving is equivalent to:

$$\begin{cases} \sum_{i:y_i=1,h_t(x)=1} H_1(X_i) > \sum_{i:y_i=1,h_t(x)=1} H_0(X_i) & ,TP \\ \sum_{i:y_i=-1,h_t(x)=-1} H_1(X_i) < \sum_{i:y_i=1,h_t(x)=1} H_0(X_i) & ,TN. \end{cases} \tag{38}$$

For AD:

$$\begin{aligned} \sum_{i:y_i=1,h_t(x)=1} H_1(X_i) &= N_{ad}\cdot\varrho\cdot 1\cdot\frac{1}{2}log\frac{1-\epsilon_t}{\epsilon_t} + N_{ad}\cdot\varrho\cdot 1\cdot z\sum_{t=1}^{T} e^{1-\epsilon_t} \\ &= \sum_{i:y_i=1,h_t(x)=1} H_0(X_i) + N_{ad}\cdot\varrho\cdot 1\cdot z\sum_{t=1}^{T} e^{1-\epsilon_t}, \end{aligned} \tag{39}$$

for z > 0, then $N_{ad}\cdot\varrho\cdot 1\cdot z\sum_{t=1}^{T} e^{1-\epsilon_t} > 0$, therefore, $\sum_{i:y_i=1,h_t(x)=1} H_1(X_i) > \sum_{i:y_i=1,h_t(x)=1} H_0(X_i)$ within matrix TP

Similarly, for non-AD samples,

$$\begin{aligned} \sum_{i:y_i=-1,h_t(x)=-1} H_1(X_i) &= \eta\cdot(-1)\cdot\frac{1}{2}log\frac{1-\epsilon_t}{\epsilon_t} + \eta\cdot(-1)\cdot z\sum_{t=1}^{T} e^{1-\epsilon_t} \\ &= \sum_{i:y_i=-1,h_t(x)=-1} H_0(X_i) + \eta\cdot(-1)\cdot z\sum_{t=1}^{T} e^{1-\epsilon_t}, \end{aligned} \tag{40}$$

where

$$\eta = N_{non-ad} - [(N_{ad} + N_{non-ad})\epsilon_t - (N_{ad} - N_{ad}\varrho)], \tag{41}$$

the last item in Eq. (40) is

$$\eta\cdot(-1)\cdot z\sum_{t=1}^{T} e^{1-\epsilon_t} = N_{ad}\cdot z\sum_{t=1}^{T}[\varrho + (\delta+1)(\epsilon_t - 1)]e^{1-\epsilon_t}. \tag{42}$$

Given that $\epsilon_t < 1\text{-}\frac{\varrho}{\delta+1}$, so $[\varrho+(\delta+1)(\epsilon_t-1)] < 0$, $\eta\cdot(-1)\cdot z\sum_{t=1}^{T} e^{1-\epsilon_t} < 0$ always hold true, therefore, $\sum_{i:y_i=-1,h_t(x)=-1} H_1(X_i) < \sum_{i:y_i=1,h_t(x)=1} H_0(X_i)$ is proved within TN metric.

Appendix C. Pseudo codes

C.1. AD multimodal data-driven adaboost Pseudo code

Algorithm 1: AD multimodal data-driven adaboost

Input:
X: Multimodal AD dataset X = $\{(x_1, y_1), \ldots, (x_N, y_N)\}$
y: Label y={-1,+1}
M: Maximum iteration
h_1: The initial diagnosis classifier.
N: Sample numbers.

Function *multimodal_adaboost(**X**,y,M,h_1,N)*
  Initialize the error weight $D^1 = \{w_i^1 | w_i^1 = \frac{1}{N}\}$ on each multimodal sample $(x_i, y_i) \in \mathbf{X}$ (i=1,...,N)
  Initialize the positive learning parameters ω and τ
  for *t =1 to M* **do**
    Train a diagnosis classifier h_t on X_t with the error weight D^t
    Calculate $z_t = \omega \cdot \mathrm{H}(X_t) + \tau \cdot \mathbb{D}(X_t)$ ▷ H and $\mathbb{D}$ are methods to calculate the entropy and variance of X_t
    Calculate the total error of h_t: $\epsilon_t = \sum_{i: y_i \neq h_t(x)} D_t(i)$
    ϵ_t should satisfy $\epsilon_t < 1 - \frac{\varrho}{\delta+1}$,else repeat this step
    Calculate the weak diagnosis classifier weight: $\alpha_1^t = \frac{1}{2} log \frac{1-\epsilon_t}{\epsilon_t} + z_t e^{1-\epsilon_t}$, where z_t is the AD modality factor
    Update $D_{t+1}(i) = D_t(i) e^{-\alpha_1^t y_i h_t(x_i)}$,i=1,...,n, renormalize it as $D_{t+1}(i) = D_{t+1}(i) / \sum_{j=1}^{n} D_{t+1}(j)$
  return $H_1(\boldsymbol{X}) = \sum_{t=1}^{T} \alpha_1^t h_t(\boldsymbol{x})$

References

Adarsh, V., Gangadharan, G., Fiore, U., & Zanetti, P. (2024). Multimodal classification of Alzheimer's disease and mild cognitive impairment using custom MKSCDDL kernel over CNN with transparent decision-making for explainable diagnosis. *Scientific Reports*, *14*(1), 1774.

Bhasin, H., Ohri, A., Kumar, N., Sharma, M., & Thakur, H. K. (2023). Mild cognitive impairment diagnosis using neuropsychological tests and agile machine learning. In *International advanced computing conference* (pp. 180–187). Springer.

Chehrehnegar, N., Nejati, V., Shati, M., Rashedi, V., Lotfi, M., Adelirad, F., et al. (2020). Early detection of cognitive disturbances in mild cognitive impairment: A systematic review of observational studies. *Psychogeriatrics*, *20*(2), 212–228.

Cortes, C., Mohri, M., & Rostamizadeh, A. (2012). L2 regularization for learning kernels. arXiv preprint arXiv:1205.2653.

De Ville, B. (2013). Decision trees. *Wiley Interdisciplinary Reviews: Computational Statistics*, *5*(6), 448–455.

El-Sappagh, S., Alonso-Moral, J. M., Abuhmed, T., Ali, F., & Bugarín-Diz, A. (2023). Trustworthy artificial intelligence in Alzheimer's disease: state of the art, opportunities, and challenges. *Artificial Intelligence Review*, *56*(10), 11149–11296.

Eskandari, A., Aghaei, M., Milimonfared, J., & Nedaei, A. (2023). A weighted ensemble learning-based autonomous fault diagnosis method for photovoltaic systems using genetic algorithm. *International Journal of Electrical Power & Energy Systems*, *144*, Article 108591.

Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., & Herrera, F. (2011). A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *42*(4), 463–484.

Gan, J., Yang, X., Zhang, G., Li, X., Liu, S., Zhang, W., et al. (2023). Alzheimer's disease pathology: pathways between chronic vascular risk factors and blood-brain barrier dysfunction in a cohort of patients with different types of dementia. *Frontiers in Aging Neuroscience*, *15*, Article 1088140.

Goel, T., Sharma, R., Tanveer, M., Suganthan, P., Maji, K., & Pilli, R. (2023). Multimodal neuroimaging based Alzheimer's disease diagnosis using evolutionary RVFL classifier. *IEEE Journal of Biomedical and Health Informatics*.

Gupta, A., & Kahali, B. (2020). Machine learning-based cognitive impairment classification with optimal combination of neuropsychological tests. In *Alzheimer's & Dementia: Translational Research & Clinical Interventions*: *vol. 6*, (1), Wiley Online Library, Article e12049.

Gustavsson, A., Norton, N., Fast, T., Frölich, L., Georges, J., Holzapfel, D., et al. (2023). Global estimates on the number of persons across the Alzheimer's disease continuum. *Alzheimer's & Dementia*, *19*(2), 658–670.

Hakan, Ö. (2024). A novel approach to detection of Alzheimer's disease from handwriting: Triple ensemble learning model. *Gazi University Journal of Science Part C: Design and Technology*, 1–1.

Hambali, M., Saheed, Y., Oladele, T., & Gbolagade, M. (2019). ADABOOST ensemble algorithms for breast cancer classification. *Journal of Advances in Computer Research*, *10*(2), 31–52.

Harciarek, M., & Cosentino, S. (2013). Language, executive function and social cognition in the diagnosis of frontotemporal dementia syndromes. *International Review of Psychiatry*, *25*(2), 178–196.

Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2006). Extreme learning machine: theory and applications. *Neurocomputing*, *70*(1–3), 489–501.

Jakkula, V. (2006). Tutorial on support vector machine (svm). *School of EECS, Washington State University*, *37*(2.5), 3.

Kandimalla, R., Saeed, M., Tyagi, N., Gupta, R. C., & Aqil, F. (2023). Exosome-based approaches in the management of Alzheimer's disease. *Neuroscience & Biobehavioral Reviews*, *144*, Article 104974.

Khan, Z., Gul, N., Faiz, N., Gul, A., Adler, W., & Lausen, B. (2021). Optimal trees selection for classification via out-of-bag assessment and sub-bagging. *IEEE Access*, *9*, 28591–28607.

Kim, S. K., Duong, Q. A., & Gahm, J. K. (2024). Multimodal 3D deep learning for early diagnosis of Alzheimer's disease. *IEEE Access*.

Kim, J., Jeong, M., Stiles, W. R., & Choi, H. S. (2022). Neuroimaging modalities in Alzheimer's disease: diagnosis and clinical features. *International Journal of Molecular Sciences*, *23*(11), 6079.

Krstev, I., Pavikjevikj, M., Toshevska, M., & Gievska, S. (2022). Multimodal data fusion for automatic detection of Alzheimer's disease. In *International conference on human-computer interaction* (pp. 79–94). Springer.

Kwon, H., Park, J., & Lee, Y. (2019). Stacking ensemble technique for classifying breast cancer. *Healthcare Informatics Research*, *25*(4), 283–288.

LaValley, M. P. (2008). Logistic regression. *Circulation*, *117*(18), 2395–2399.

Lei, B., Hou, W., Zou, W., Li, X., Zhang, C., & Wang, T. (2019). Longitudinal score prediction for Alzheimer's disease based on ensemble correntropy and spatial–temporal constraint. *Brain Imaging and Behavior*, *13*, 126–137.

Leon, F., Floria, S. A., & Bădică, C. (2017). Evaluating the effect of voting methods on ensemble-based classification. In *2017 IEEE international conference on iNnovations in intelligent systems and applications* (pp. 1–6). IEEE.

Li, W., Zhao, J., Shen, C., Zhang, J., Hu, J., Xiao, M., et al. (2022). Regional brain fusion: Graph convolutional network for Alzheimer's disease prediction and analysis. *Frontiers in Neuroinformatics*, *16*, Article 886365.

Lin, L., Xiong, M., Zhang, G., Kang, W., Sun, S., Wu, S., et al. (2023). A convolutional neural network and graph convolutional network based framework for AD classification. *Sensors*, *23*(4), 1914.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., et al. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 10012–10022).

Loewenstein, D. A., Acevedo, A., Agron, J., Issacson, R., Strauman, S., Crocco, E., et al. (2006). Cognitive profiles in Alzheimer's disease and in mild cognitive impairment of different etiologies. *Dementia and Geriatric Cognitive Disorders*, *21*(5–6), 309–315.

Malek-Ahmadi, M., Davis, K., Belden, C. M., & Sabbagh, M. N. (2014). Comparative analysis of the Alzheimer questionnaire (AQ) with the CDR sum of boxes, MoCA, and MMSE. *Alzheimer Disease & Associated Disorders*, *28*(3), 296–298.

Markwick, A., Zamboni, G., & de Jager, C. A. (2012). Profiles of cognitive subtest impairment in the montreal cognitive assessment (MoCA) in a research cohort with normal mini-mental state examination (MMSE) scores. *Journal of Clinical and Experimental Neuropsychology*, *34*(7), 750–757.

Mathew, T. E., Kumar, K. A., & Kumar, K. S. (2020). Breast cancer diagnosis using stacking and voting ensemble models with Bayesian methods as base classifiers. *IX*, (Ii), (pp. 108–121).

McKee, C. G., Hoffos, M., Vecchiarelli, H. A., & Tremblay, M.-È. (2023). Microglia: A pharmacological target for the treatment of age-related cognitive decline and Alzheimer's disease. *Frontiers in Pharmacology*, *14*, Article 1125982.

Pan, Z., & Mishra, P. (2023). Hardware trojan detection using Shapley ensemble boosting. In *Proceedings of the 28th Asia and south Pacific design automation conference* (pp. 496–503).

Paplomatas, P., Krokidis, M. G., Vlamos, P., & Vrahatis, A. G. (2023). An ensemble feature selection approach for analysis and modeling of transcriptome data in Alzheimer's disease. *Applied Sciences*, *13*(4), 2353.

Prodromidis, A. L., & Stolfo, S. J. (1999). A comparative evaluation of meta-learning strategies over large and distributed data sets. In *Workshop on meta-learning, sixteenth intl. conf. machine learning* (pp. 18–27).

Rahim, N., Abuhmed, T., Mirjalili, S., El-Sappagh, S., & Muhammad, K. (2023). Time-series visual explainability for Alzheimer's disease progression detection for smart healthcare. *Alexandria Engineering Journal*, *82*, 484–502.

Rahim, N., El-Sappagh, S., Ali, S., Muhammad, K., Del Ser, J., & Abuhmed, T. (2023). Prediction of Alzheimer's progression based on multimodal deep-learning-based fusion and visual explainability of time-series data. *Information Fusion*, *92*, 363–388.

Rayana, S., Zhong, W., & Akoglu, L. (2016). Sequential ensemble learning for outlier detection: A bias-variance perspective. In *2016 IEEE 16th international conference on data mining* (pp. 1167–1172). IEEE.

Ribeiro, D. E., Petiz, L. L., Glaser, T., Oliveira-Giacomelli, Á., Andrejew, R., Saab, F. d. R., et al. (2023). Purinergic signaling in cognitive impairment and neuropsychiatric symptoms of Alzheimer's disease. *Neuropharmacology*, *226*, Article 109371.

Schmidt, M., Fung, G., & Rosales, R. (2009). *Optimization methods for l1-regularization*: *Technical Report TR-2009-19*, University of British Columbia.

Shafiei, A., Tatar, A., Rayhani, M., Kairat, M., & Askarova, I. (2022). Artificial neural network, support vector machine, decision tree, random forest, and committee machine intelligent system help to improve performance prediction of low salinity water injection in carbonate oil reservoirs. *Journal of Petroleum Science and Engineering*, *219*, Article 111046.

Shukla, A., Tiwari, R., & Tiwari, S. (2024). Analyzing subcortical structures in Alzheimer's disease using ensemble learning. *Biomedical Signal Processing and Control*, *87*, Article 105407.

Song, X., Mao, M., & Qian, X. (2021). Auto-metric graph neural network based on a meta-learning strategy for the diagnosis of Alzheimer's disease. *IEEE Journal of Biomedical and Health Informatics*, *25*(8), 3141–3152.

Song, X., Zhou, F., Frangi, A. F., Cao, J., Xiao, X., Lei, Y., et al. (2022). Multicenter and multichannel pooling GCN for early AD diagnosis based on dual-modality fused brain network. *IEEE Transactions on Medical Imaging*, *42*(2), 354–367.

Tosi, G., Borsani, C., Castiglioni, S., Daini, R., Franceschi, M., & Romano, D. (2020). Complexity in neuropsychological assessments of cognitive impairment: A network analysis approach. *Cortex*, *124*, 85–96.

Ying, Y., Yang, T., & Zhou, H. (2023). Multimodal fusion for Alzheimer's disease recognition. *Applied Intelligence: The International Journal of Artificial Intelligence, Neural Networks, and Complex Problem-Solving Technologies*, *53*(12), 16029–16040.

Zhang, M., Cui, Q., Lü, Y., & Li, W. (2024). A feature-aware multimodal framework with auto-fusion for Alzheimer's disease diagnosis. *Computers in Biology and Medicine*, *178*, Article 108740.

Zhang, J., He, X., Liu, Y., Cai, Q., Chen, H., & Qing, L. (2023). Multi-modal cross-attention network for Alzheimer's disease diagnosis with multi-modality data. *Computers in Biology and Medicine*, *162*, Article 107050.

Zhang, Y., Shen, Y., Liufu, N., Liu, L., Li, W., Shi, Z., et al. (2023). Transmission of Alzheimer's disease-associated microbiota dysbiosis and its impact on cognitive function: evidence from mice and patients. *Molecular Psychiatry*, *28*(10), 4421–4437.

Zhu, Z., Lu, S., Wang, S. H., Gorriz, J. M., & Zhang, Y. D. (2022). DSNN: A DenseNet-based SNN for explainable brain disease classification. *Frontiers in Systems Neuroscience*, *16*, Article 838822.