Classification and prediction of Alzheimer's disease based on machine learning KAN network

Wenxin Xu[#, *], Sihan Long[#], Jiayu Shi[#]

Faculty of Science and Technology, University of Macau, Macau, China, 999078

*Corresponding author: dc12721@um.edu.mo

#These authors contributed equally.

Abstract. Alzheimer's disease can cause memory and cognitive decline, but early and accurate diagnosis is not easy. With the continuous enrichment of neuroimaging data, the progress of genetics, biochemistry, and other disciplines has made it possible to study Alzheimer's disease (AD) in multiple dimensions. The ever-evolving machine learning and deep learning algorithms overcome the shortcomings of high dimensionality and small sample size of neuroimaging, and provide advanced models for clinical intelligent non-invasive diagnosis. This article reviews the application of machine learning methods and strategies based on the ADNI database in the early diagnosis of mild cognitive dysfunction and AD, and uses random forest, support vector machine, multilayer perceptron (MLP), and Kolmogorov-Arnold Networks (KAN) to predict and classify the early clinical diagnosis of Alzheimer's disease. The high accuracy of KAN network indicates that the machine learning method is helpful to improve the accuracy of predicting AD subtypes.

Keywords: Alzheimer's disease; Random forest; Confusion matrix; Bayesian optimization.

1. Introduction

Alzheimer's disease (AD) is a progressive and degenerative form of dementia characterized by symptoms such as memory loss, cognitive decline, speech impairment, and disorientation, which worsen over time and eventually lead to death [1]. Clinically, AD manifests as a decline in learning abilities, memory impairment, partial loss of language function, and reduced mobility, significantly impacting patients' physical health and quality of life, as well as placing a heavy burden on their families [2]. Dementia is particularly prevalent among the elderly. Statistics reveal that approximately 9% of individuals aged over 65 are affected by dementia, and this figure rises dramatically to 34% among those aged over 85. With global populations aging rapidly, the prevalence of Alzheimer's disease is anticipated to increase substantially in the coming decades.

Currently, there is no cure or definitive treatment for AD. Available interventions focus primarily on symptom management and slowing disease progression. Early detection is critical; timely diagnosis enables healthcare professionals to develop individualized care plans that can delay the disease's advancement and improve the patient's quality of life. Consequently, designing high-precision computer-aided diagnosis (CAD) systems for Alzheimer's disease has become an area of intense research interest. These systems aim to enhance diagnostic accuracy, optimize treatment strategies, and reduce diagnostic delays.

Mild cognitive impairment (MCI) is recognized as an intermediate stage between normal aging and Alzheimer's disease, often referred to as the preclinical phase of AD [3]. Early identification and intervention in MCI are crucial, as studies have demonstrated that appropriate pharmacological treatments, combined with advanced medical technologies and therapeutic strategies, can significantly alleviate symptoms. In some cases, effective interventions have even enabled MCI patients to return to a near-normal cognitive state [4]. Early diagnosis of MCI not only offers an opportunity to halt or reverse cognitive decline but also provides a critical window for implementing preventive measures, reducing the progression to full-blown Alzheimer's disease. As a result, developing accurate and efficient diagnostic tools for MCI has become a focal point for researchers and clinicians seeking to combat the growing prevalence of neurodegenerative disorders.

This study aims to address these challenges by developing a high-precision CAD system that leverages advanced artificial intelligence algorithms and integrates multimodal data such as brain imaging, biomarkers, and cognitive test results. By doing so, the research seeks to enhance diagnostic accuracy, provide personalized intervention strategies, and reduce diagnostic delays. The significance of this work lies in its potential to advance scientific understanding of neurodegenerative diseases, improve clinical practices by enabling early and precise interventions, and alleviate the societal and economic burdens associated with AD. Key innovations include applying cutting-edge deep learning techniques optimized for small medical datasets, integrating multimodal data for comprehensive analysis, and the development of personalized risk prediction models to guide tailored treatment plans. Moreover, the system is designed for practical use in clinical settings and is scalable, offering a generalized framework for diagnosing and managing neurodegenerative disorders effectively.

2. Basic function

2.1. Data Sources and Preparation

2.1.1 Datasets

The data used in this article are from the ADNI (Alzheimer's Disease Neuroimaging Initiative) database, which is roughly divided into three categories according to the severity of the patient's condition: Normal Control (NC), Mild Cognitive Impairment (MCI), and Alzheimer's Disease (AD), with 220, 402, and 192 data, respectively. For MCI, it can be further divided into light MCI (lMCI) and stable MCI (sMCI) groups, which have 146 and 256 data, respectively. At this time, there are four types of data labels: NC, lMCI, sMCI, and AD, and four classification tasks can be performed at this time.

2.1.2 Data preprocessing

After the missing values and outliers of the original data are processed, the DX_bl data is encoded with tags such as "0" for "CN", "1" for SMC, "2" for EMCI, "3" for LMCI, and "4" for AD. Race, skin color, marital status, and gender are coded using unique data tags. The data were z-score normalized so that the variance of different variables was consistent.

2.2. Experimental environment

This study utilizes Pycharm software and Python for the experiments. The SVM model is implemented using the libsvm open-source toolbox (version 2.91), with model performance evaluated through 5-fold cross-validation. The penalty parameter C and kernel function G for the SVM classifier are selected via grid search within the range of $\{2^{-10}, 2^{-9}, ..., 2^{9}, 2^{10}\}$. For the MLP and KAN networks, parameter selection is conducted using the Optuna library for Bayesian hyperparameter optimization, employing the Tree-structured Parzen Estimator (TPE) algorithm to automate the optimization process. This approach allows for flexible modeling of various factors and better handling of missing data and noise. The evaluation metrics for the model include accuracy (Acc), precision (Pre), and f1-score (F1). Accuracy is the proportion of correct predictions to all predictions, and is calculated as follows:

$$Acc = \frac{TP+TN}{TP+FN+FP+TN} \tag{1}$$

The precision Pre is the proportion of all samples with positive predictions that are correctly predicted, and is calculated as follows:

$$Pre = \frac{TP}{TP+FP} \tag{2}$$

F1 measures P and R in combination and is calculated as the harmonic mean of the two, which is calculated as follows:

$$F1 = \frac{2TP}{2TP+FP+FN} \tag{3}$$

Where TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative examples, respectively.

In this experiment, accuracy, weighted precision, and weighted f1-scores are used to evaluate the performance of our model and are represented by Acc, Weight_Pre, and Weight_F1, respectively.

2.3. Experimental methods

2.3.1 Random Forest Classifier

Random forest is an algorithm that integrates multiple trees through the bagging idea of ensemble learning. It builds multiple decision trees by randomly selecting samples and features, and determines the final outcome using a "vote" or averaging method. It can be used for classification and regression problems, and random forests generally have better generalization performance than single decision trees.

2.3.2 Support vector machine (SVM)

SVM is a supervised learning algorithm for binary and multi-class classification. It identifies an optimal hyperplane in the feature space that maximizes the margin, employing structural risk minimization and interval maximization strategies. By relying on a small number of support vectors, SVM achieves a globally optimal solution with strong generalization capabilities, avoiding local optima common in neural networks. Its computational complexity depends on the number of support vectors rather than the sample space's dimensionality, mitigating the "curse of dimensionality." SVM also handles non-linear data effectively through kernel functions and relaxation variables, making it robust against overfitting and suitable for large datasets, such as those involving Alzheimer's symptoms.

2.3.3 Multilayer perception (MLP)

MLP is a type of feedforward neural network where information flows from the input layer, through hidden layers, to the output layer [5]. Comprising interconnected neurons, each performs a weighted sum of inputs, applies an activation function, and passes the result forward. The hidden layers allow MLPs to learn complex data representations, making them fundamental for approximating nonlinear functions, as guaranteed by the universal approximation theorem. MLPs are widely used in supervised learning tasks like classification.

2.3.4 Kolmogorov-Arnold Networks (KAN)

MLPs are rooted in the universal approximation theorem, while KANs draw on the Kolmogorov-Arnold representation theorem, offering a flexible alternative [6]. Unlike MLPs with fixed activation functions, KANs use learnable activation functions on edges, replacing linear weights with univariate spline functions. Known for accuracy and compactness, KANs reduce computational demands and improve training efficiency while maintaining strong generalization [7]. They adapt dynamically to different datasets and tasks, enabling effective nonlinear mapping in high-dimensional spaces. This study explores the application of multi-layer KANs for multi-classification problems, integrating a scheduler and different optimizers to enhance real-world adaptability.

3. Result and Analysis

3.1. Random Forest Classifier

3.1.1 Feature Selection

Use the Random Forest Classifier's built-in attribute feature_importances to extract the importance value of each feature in the model. After extracting the importance value, filter out the variables with non-zero importance values. The variables with 0 feature importance are regarded as redundant features, that is, they have no independent contribution to the decision-making process of the model, and the remaining 19 features and importance scores are shown in the following Table 1:

Table 1: Nineteen feature and importance scores

Variable	Importance
CDRSB_bl	0.22
MMSE_bl	0.08
ADAS13_bl	0.07
FAQ_bl	0.07
ADAS11_bl	0.06
RAVLT_immediate_bl	0.05
Hippocampus_bl	0.04
Entorhinal_bl	0.04
Fusiform_bl	0.04
MidTemp_bl	0.04
ICV_bl	0.04
AGE	0.04
RAVLT_learning_bl	0.03
RAVLT_perc_forgetting_bl	0.03
Ventricles_bl	0.03
WholeBrain_bl	0.03
RAVLT_forgetting_bl	0.02
PTEDUCAT	0.02
APOE4	0.01

3.1.2 Random forest classifier tuning parameters

After training a random forest classifier to predict the severity of Alzheimer's disease in the test set, the classifier achieved an accuracy of 78.9%. The optuna library was used for tuning, so the accuracy was kept at the ideal value.

Table 2: Optuna Library Performs Tuning Scores for random forests

Name	Value
n_estimators	6
min_samples_split	4
min_samples_leaf	3
min_weight_fraction_leaf	0.0008370
max_features	14
min_impurity_decrease	0.008342
best score	0.7610

3.1.3 Confusion matrix for random forests

The classification performance of random forests is evaluated by a confusion matrix, where each element in the diagonal of the confusion matrix represents the classification of the predicted results, where the elements on the diagonal represent the number of correct classifications, and the elements on the non-diagonal lines represent the number of misclassifications. It is worth noting that the diagonal element corresponding to the SMC in the random forest is 0, indicating that the model did not correctly predict any samples of the SMC. This may be due to the very small sample size of the SMC, which appears only 234 times out of 1347 pieces of data, so the model may have difficulty learning the features of the SMC and thus cannot correctly classify the samples of that category in the test set. The confusion matrix of the random forest classifier are shown in Figure 1:

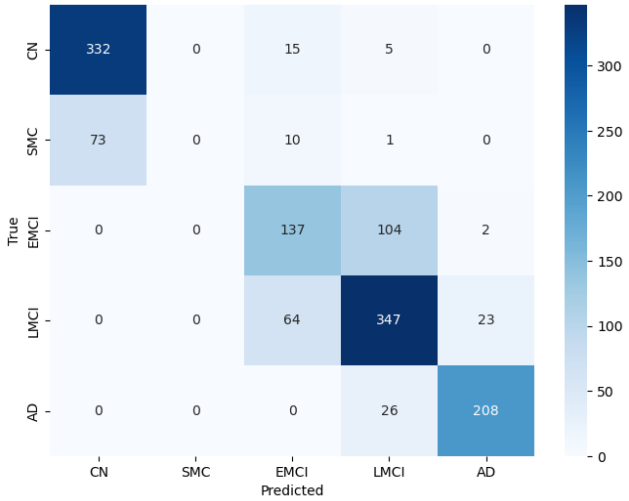

Figure 1: Combined confusion matrix of random forest

3.1.4 Random forest classifier performance analysis

The accuracy of the random forests model fluctuated between 0.7546 and 0.7695, with an average accuracy of 0.7602, and the overall performance was stable. The weighted accuracy fluctuated from 0.7040 to 0.7191, and the average weighted accuracy was 0.7109, indicating that the accuracy of the model was average in the case of unbalanced processing of samples. The weighted f1 score fluctuated between 0.7268 and 0.7419, and the average weighted f1 score was 0.7330, which comprehensively considered the accuracy and recall of the model, and the overall performance was good. The results of the random forest model are shown in Table 3:

Table 3: Accuracy, weighted precision, weighted F1 score of random forest

Name	1st	2nd	3rd	4th	5th	Mean
Acc	0.7593	0.7556	0.7695	0.7621	0.7546	0.7602
Weight_Pre	0.7069	0.7040	0.7191	0.7142	0.7102	0.7109
Weight_F1	0.7310	0.7268	0.7419	0.7369	0.7281	0.7330

The model performed relatively well in accuracy and weighted f1 score, but it was slightly insufficient in weighted accuracy, and its overall performance was stable, indicating that the model had certain stability and reliability.

3.2. Support vector machine

3.2.1 Confusion matrix for SVM

The performance of the Support Vector Machine (SVM) model is evaluated by an obfuscation matrix, where each element on the diagonal represents the number of correctly classified, and the non-diagonal element represents the number of misclassified. As with the random forest classifier, the SVM model has a diagonal element of 0 for the SMC category, indicating that the model did not correctly predict any samples of the SMC. The possible reason is that the sample size of the SMC is very small, only 234 times out of 1347 pieces of data, so the model may have difficulty learning the features of the SMC, so that the samples of this category cannot be correctly classified in the test set. The confusion matrix of the SVM model are shown in Figure 2:

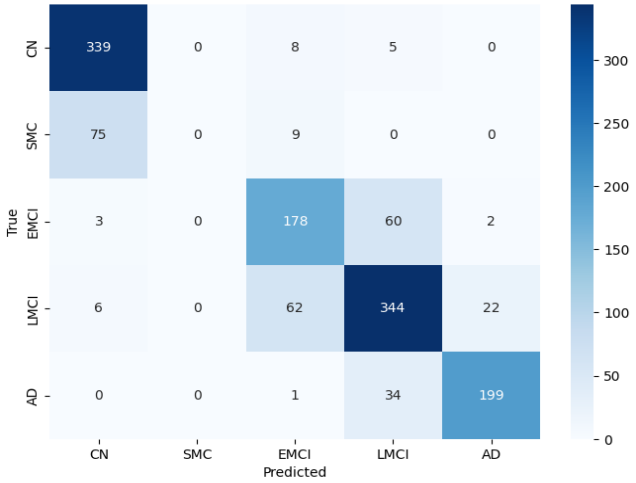

Figure 2: Combined confusion matrix of SVM

3.2.2 Performance analysis of the SVM model

The accuracy of the SVM model fluctuated from 0.7770 to 0.7963, and the average accuracy of the five experiments was 0.7869, and the overall performance of the model was relatively stable and higher than the average accuracy of the random forest model. The weighted precision fluctuates between 0.7358 and 0.7525, and the average weighted precision is 0.7411, which is slightly higher than the average weighted accuracy of the random forest model, but still not satisfactory. The weighted f1 score fluctuated from 0.7508 to 0.7690, and the average weighted f1 score was 0.7607, which was slightly higher than the average weighted f1 score of the random forest model, showing the relative advantage of the model when considering accuracy and recall.

SVMs perform relatively well on accuracy and weighted f1 scores, especially in accuracy.

Although slightly higher than the random forest model in terms of weighted precision, there is still room for improvement, especially when dealing with unbalanced datasets, which may require more tweaking and optimization. The results of the SVM model are shown in Table 4:

Table 4: SVM's accuracy, weighted precision, weighted F1 score, and average

Name	1st	2nd	3rd	4th	5th	Mean
Acc	0.7963	0.7815	0.7844	0.7955	0.7770	0.7869
Weight_Pre	0.7462	0.7358	0.7381	0.7525	0.7331	0.7411
Weight_F1	0.7690	0.7561	0.7585	0.7693	0.7508	0.7607

The average performance of the SVM model is stable and fluctuates in a small range, indicating that the model performs relatively consistently across different datasets or cross-validation trade-offs. This dataset shows high accuracy and weighted F1 scores, but there is room for improvement in weighted accuracy. The SVM classifier has some advantages for handling more complex classification problems, but it still needs to be further optimized.

3.3. Multilayer Perceptron Classifier (MLP)

3.3.1 Multilayer Perceptron Classifier Tuning Parameters

After training a multi-layer perceptron classifier to predict the severity of Alzheimer's disease in the test set, the classifier achieved an accuracy of 78.1%. The optuna library was used for tuning, so the accuracy was kept at the ideal value. The results of the MLP Classifier are shown in Table 5:

Table 5: Parameter settings for the Multi-layer Perceptron Classifier

Name	Value
Solver	Adam
Learning_rate_init	0.003442
Activation	Relu

3.3.2 Multilayer Perceptron Classifier Confusion Matrix

The performance of the MLP model is evaluated by a confusion matrix, where each element on the diagonal represents the number of correct classifications, and the non-diagonal elements represent the number of misclassifications. Unlike the random forest and the SVM model, MLP has a diagonal element of 5 for the SMC category, indicating that the model can correctly predict some SMC samples. However, the SMC sample size is still small, only 234 occurrences in 1347 pieces of data, so it may be difficult for the model to fully learn the features of SMC, so the classification effect in the test set is limited. The MLP confusion matrix results are shown in Figure 3 below:

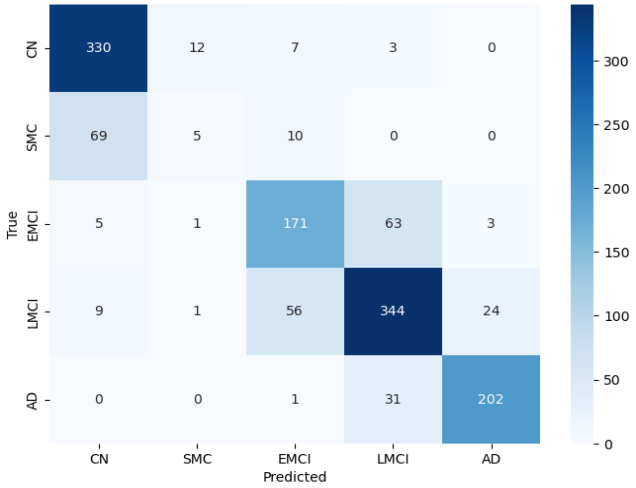

Figure 3: Combined confusion matrix of MLP

3.3.3 Performance analysis of Multilayer Perceptron Classifier

The accuracy fluctuated between 0.7695 and 0.7963, averaging 0.7810. This indicates that MLP is more stable in different experiments and is able to classify correctly in most cases. The weighted precision ranged from 0.7256 to 0.8081, with an average of 0.7627. This reflects the effectiveness of the model when dealing with different classes, but there may be category imbalance effects. The weighted F1 score ranged from 0.7465 to 0.7779, with an average of 0.7625, showing a good trade-off between accuracy and recall. The results of MLP are shown in the Table 6 below:

Table 6: Accuracy, weighted precision, weighted F1 score, and mean of MLP

Name	1st	2nd	3rd	4th	5th	Mean
Acc	0.7963	0.7704	0.7732	0.7955	0.7695	0.7810
Weight_Pre	0.8081	0.7256	0.7506	0.7747	0.7545	0.7627
Weight_F1	0.7732	0.7465	0.7584	0.7779	0.7567	0.7625

The MLP model showed good classification ability in this set of experiments, especially when dealing with complex data, and was able to maintain high accuracy and precision. Although the accuracy of some experiments is slightly lower, the overall performance is well-balanced in all indicators. This shows that MLP is suitable for classification tasks that require stability and accuracy. To further improve performance, consider data augmentation or tuning model parameters to improve minority identification.

3.4. Kolmogorov-Arnold Networks (KANs)

3.4.1 Kolmogorov-Arnold Networks Tuning Parameters

After training a Kolmogorov-Arnold Networks to predict the severity of Alzheimer's disease in the test set, the classifier achieved an accuracy of 87.6%. The optuna library was used for tuning, so the accuracy was kept at the ideal value. The results of KAN are shown in Table 7 below:

Table 7: Kolmogorov-Arnold Networks parameter settings

Name	Value
Solver	Adam
Learning_rate_init	0.004276
Activation	Relu

3.4.2 Kolmogorov-Arnold Networks confusion matrix

The performance of KAN is evaluated by a confusion matrix, where each element on the diagonal represents the value of the correct classification: CN: 326, SMC: 44, EMCI: 19, LMCI: 387, AD: 223. In contrast, the previous MLP model performed poorly on the SMC category at only 5, while both the random forest and SVM had 0 diagonal elements for the SMC category, showing an inability to correctly predict SMC samples. KAN was able to achieve a correct classification of 44 on the SMC category, indicating that it has a significant advantage in capturing minority features.

Kolmogorov-Arnold Networks achieved good classification results in SMC categories, which was significantly better than random forest, SVM and MLP. This suggests that KAN is more advantageous when dealing with classes with small sample sizes and complex features. In terms of the overall performance of the model, KAN also demonstrated good classification capabilities for other categories (CN, EMCI, LMCI, AD), especially for LMCI and AD categories.

Compared with random forest, SVM and MLP, KAN has better classification performance, especially in dealing with data imbalance and minority problems. This may be attributed to the fact that the structure of the KAN is better able to adapt to the complexity and diversity of the data. With further optimization, KAN has the potential to achieve greater accuracy and reliability in a variety of classification tasks.

3.4.3 Kolmogorov-Arnold Networks Performance Analysis

For the results of the original KANs model, it was able to maintain a high level of accuracy in most of the experiments, especially in the last few experiments. The weighted precision achieves an average of 0.8747. This indicates that KAN has good accuracy when processing different types of samples, and can effectively distinguish between different types of data, with a weighted f1-score of KAN reaching an average of 0.8730, indicating a good balance between accuracy and recall compared with the other models mentioned above. The KAN results of weighted weight are shown in Table 8:

Table 8: Accuracy, weighted precision, weighted F1 score of KAN

Name	1st	2nd	3rd	4th	5th	Mean
Acc	0.8622	0.8778	0.8773	0.8691	0.8905	0.8754
Weight_Pre	0.8604	0.8767	0.8750	0.8714	0.8900	0.8747
Weight_F1	0.8605	0.8759	0.8725	0.8663	0.8897	0.8730

Kolmogorov-Arnold Networks has demonstrated strong performance on these metrics, especially in tasks with high complexity and data diversity. Its high average accuracy and precision show that KAN can process data efficiently and has strong ability to learn minority features. This makes KAN a reliable choice for a variety of classification tasks, especially in scenarios that require high generalization capabilities. With further optimization, KAN has the potential to achieve superior performance in a wider range of applications.

3.4.4 KAN performance analysis tuned by scheduler and R2 regularization

Based on the comparison of the random forest, SVM, MLP, and KAN models, the KAN model was selected for its superior performance. To further optimize the KAN model, techniques such as scheduler and L2 regularization on weights were applied.

In this experiment, cosine annealing was chosen to be the scheduler, which dynamically adjusts the learning rate during training to improve convergence and avoid local optima or oscillations [8]. Additionally, weight decay played a role in implementing L2 regularization that adds penalty terms

to model weights during updates to prevent overfitting and improve generalization [9]. The combined use of cosine annealing and weight decay significantly improves the KAN model's stability, generalization ability, and resistance to overfitting. The modified KAN model demonstrated superior performance with an average accuracy of 88.36%, outperforming the original KAN model.

4. Conclusion

This study focused on developing a machine learning-based approach for the automatic classification of Alzheimer's disease (AD), addressing challenges such as small sample sizes and high feature dimensions in MRI data. By employing a low-rank learning method for feature selection and optimizing the KAN network using a simulated annealing approach, the study achieved significant advancements in diagnostic accuracy. The modified KAN model demonstrated superior performance with an accuracy of 88.36%, outperforming other classifiers such as random forest, SVM, and MLP. The results underscore the efficacy of multi-template features compared to single-template features in improving classification performance.

These findings highlight the potential of artificial intelligence (AI)-driven methods in addressing limitations of traditional diagnostic approaches, particularly for small-sample datasets. Deep learning strategies, including transfer learning and residual networks, have shown promise in enhancing prediction accuracy and overcoming challenges in MRI-based AD classification. Future research should focus on expanding datasets through collaboration, exploring advanced multimodal data integration techniques, and refining AI algorithms to handle individual variability and subtle differences in brain structures. Additionally, combining AI-based tools with established diagnostic methods like PET imaging and cerebrospinal fluid analysis could further enhance early detection and intervention capabilities for Alzheimer's disease.

Reference

[1] Mckhann G, Drachman D, Folstein M, et al. Clinical diagnosis of Alzheimer’s disease Report of the NINCDS-ADRDA work group under the auspices of department of health and human services task force on Alzheimer’s disease [J]. Neurology, 1984, 34 (7): 939-939.

[2] Ott A, Breteler M, Van H F, et al. Prevalence of Alzheimer’ s disease and vascular dementia: Association with education. the rotterdam study [J]. BMJ, 1995, 310 (6985): 970-973.

[3] LIU H Y, XIA Z Y, SHI Y, et al. Overview of pathogenesis of Alzheimer’s disease [J]. Geriatrics & Health Care, 2020, 26(6): 1098-1100.

[4] ZHANG W Y, LU Y. Advances in etiology and treatment of mild cognitive impairment in older people [J]. International Journal of Digestive Diseases, 2021, 42(1): 57-61.

[5] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. Neural networks, 2(5):359–366, 1989.

[6] LIU Z, WANG Y, Vaidya S, et al. Kan: Kolmogorov-arnold networks[J]. arXiv preprint arXiv: 2404. 19756, 2024.

[7] ZHANG B W, LIN L, WU S C. Efficient Alzheimer’s disease classification based on AlexNet model [J]. Journal of Beijing University of Technology, 2020, 46(1): 68-74.

[8] Basheera S, Ram M S S. A novel CNN based Alzheimer’s disease classification using Hybrid enhanced ICA segmented gray matter of MRI [J]. Computerized Medical Imaging and Graphics, 2020, 81: 101713.

[9] PAN D, ZENG A, JIA L, et al. Early Detection of Alzheimer's Disease Using Magnetic Resonance Imaging: A Novel Approach Combining Convolutional Neural Networks and Ensemble Learning[J]. Front Neurosci, 2020, 14: 259.

[10] Y.H. He. Machine Learning in Pure Mathematics and Theoretical Physics. G- Reference, Information and Interdisciplinary Subjects Series[J]. World Scientific, 2023.