Enhancing 3D Multimodal Image Classification via Multi-plane View with Ensemble Multi-tree Genetic Programming for Alzheimer's Disease Diagnosis

Yuye Zhang, Fangfang Zhang✉, Bing Xue, Mengjie Zhang
Centre for Data Science and Artificial Intelligence & School of Engineering and Computer Science
Victoria University of Wellington, PO Box 600, Wellington, New Zealand
Email: {yuye.zhang, fangfang.zhang, bing.xue, mengjie.zhang}@ecs.vuw.ac.nz

***Abstract*—Multimodal neuroimaging using magnetic resonance imaging (MRI) and positron emission tomography (PET) provides complementary structural and functional information for Alzheimer's disease (AD) diagnosis. However, effective classification of high-dimensional 3D multimodal data remains challenging due to the limited number of samples and high computational complexity. This paper proposes a 3D ensemble multimodal multi-plane multi-tree genetic programming (3DMPGP) method for AD diagnosis using 3D MRI and PET data. The proposed approach employs a multi-tree GP representation to learn from 3D images across three anatomical planes—axial, sagittal, and coronal. Experimental results on the Australian Imaging, and Biomarker and Lifestyle Flagship Study of Aging (AIBL) demonstrate that 3DMPGP consistently outperforms competitive GP-based ensemble baselines, as well as representative neural network based methods. These results and further analysis indicate that the proposed method effectively explores complementary multimodal information and comprehensive multi-plane characteristics, enhancing 3D image understanding from different anatomical views.**

***Index Terms*—Genetic programming, multimodal learning, ensemble learning, 3D medical image analysis, Alzheimer's disease diagnosis**

I. Introduction

Alzheimer's disease (AD) is one of the most prevalent neurodegenerative disorders and poses a growing challenge to public health worldwide, which means accurate diagnosis is critical for timely intervention and disease management [1]. Neuroimaging has become a popular and essential tool for supporting AD diagnosis, as it enables non-invasive assessment of brain structure and function. Magnetic resonance imaging (MRI) provides high-resolution three-dimensional (3D) structural information, while positron emission tomography (PET) produces lower-resolution but highly informative 3D functional measurements related to metabolic or pathological changes [2]. Thus, the MRI and PET modalities provide complementary structural and functional information, enabling a better characterization of disease-related brain alterations than using a single modality.

Despite their high diagnostic value, effective analysis of multimodal neuroimaging data in full 3D form remains challenging. First, 3D data contain orders of magnitude more voxels than 2D data, leading to substantially higher computational complexity and cost when processing full volumetric images. Second, multimodal data sources further increase modeling complexity, as MRI and PET have different characteristics. Moreover, the relatively limited number of samples typically available in AD studies further increases the risk of overfitting.

Existing approaches proposed for AD diagnosis exhibit several limitations when applied to multimodal 3D neuroimaging data. Although deep learning approaches have become increasingly popular for neuroimaging analysis, their complex and black-box model structures make the underlying decision-making process difficult for clinicians and patients to understand [3]. Such limited transparency can prevent clinical trust and acceptance, particularly in medical diagnosis tasks. Classical machine learning methods rely heavily on pre-extracted features and often process two-dimensional images on a predefined plane view [4], making it difficult to fully use the spatial information present in 3D images. Genetic programming (GP) has been explored for image classification due to its flexible representation and interpretability [5], but most existing GP-based approaches focus on input data on a single plane. A plane here refers to a 2D cross-sectional view from a 3D image. Specifically, for 3D brain scans, the three planes are the axial (top–bottom), sagittal (left–right), and coronal (front–back) planes defined by anatomy.

However, existing GP-based methods for medical image classification are largely restricted to single-plane learning and therefore fail to understand the 3D neuroimaging data from the three-dimensional structure. The use of multi-plane view is motivated by the inherently three-dimensional nature of neuroimaging data, where disease-related patterns often exist across multiple adjacent slices. Relying on 2D slices from a single anatomical plane is therefore insufficient to fully capture such spatial characteristics, as important inter-slice and cross-plane information may be lost. Incorporating multiple anatomical planes enables a more comprehensive characterization of structural and functional changes associated with disease.

To address the limitations, this paper proposes a 3D ensemble multimodal multi-plane multi-tree Genetic Programming (3DMPGP) method for Alzheimer's disease diagnosis. The contributions of this paper are summarized as follows:

- We propose a multi-plane voting strategy that models 3D neuroimaging data through key slices in axial, sagittal, and coronal anatomical planes, addressing the difficulty

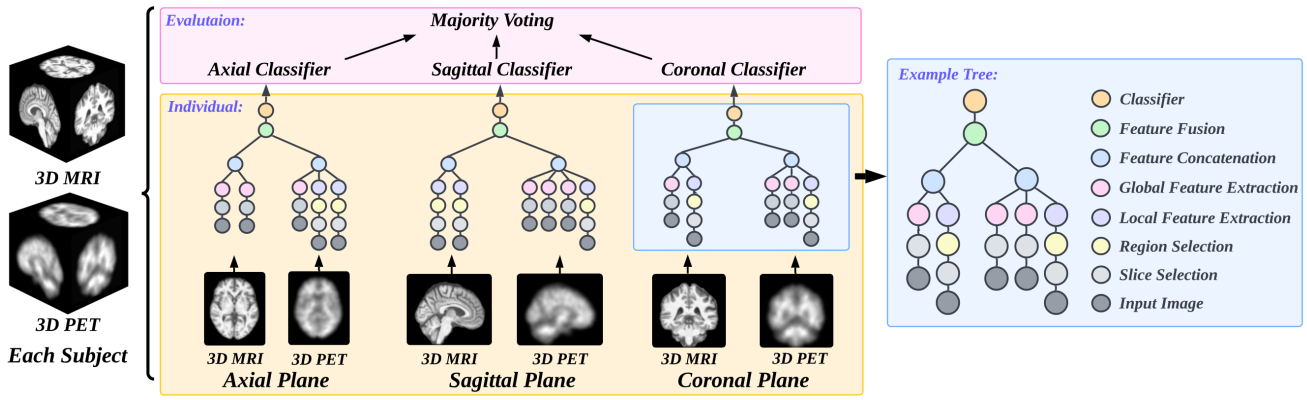

Fig. 1. Outline of the proposed 3DMPGP approach using multi-tree representation.

of directly learning discriminative patterns from high-dimensional 3D images. This design effectively captures comprehensive information via multi-plane views and improves robustness and performance in 3D Alzheimer's disease classification.

- We develop a multi-tree genetic programming representation to model 3D images from multiple anatomical planes, where each tree performs modality-specific feature learning and multimodal fusion within a single plane, addressing the limitation of relying on a single modality. This structured within-tree fusion enables flexible exploration of complementary MRI and PET information, leading to improved classification performance over using a single modality.
- We employ a frequency-based strategy to visualize the brain regions selected from different anatomical planes. This analysis highlights AD-related abnormal regions identified by the model, and the consistently selected regions align well with known anatomical patterns associated with Alzheimer's disease, providing clinically meaningful interpretability.

II. Background and Related Work

Multimodal neuroimaging, particularly the combination of MRI and PET, has been widely investigated for AD diagnosis, as these modalities provide complementary anatomical and functional information. Early studies mainly relied on pre-extracted features with traditional machine learning models for classification [4]. However, these methods heavily depend on manual feature design and separate learning stages, which limit their ability to capture complex and discriminative patterns from 3D neuroimaging data. More recently, deep learning–based methods have demonstrated promising capability by learning directly from 3D data [6]. However, unlike traditional machine learning models that produce clear decision boundaries, deep neural networks have been criticized for acting as complex "black boxes", making it hard to understand the logic behind the AD predictions [3]. In addition, designing suitable network architectures and tuning hyperparameters usually involves balancing model complexity and parameter size, which is challenging under the limited data and computational constraints common in AD studies [7].

GP has been increasingly used for image classification tasks due to its ability to evolve flexible and interpretable solutions. Recent studies have demonstrated the effectiveness of GP in medical tasks, particularly in medical imaging. To enhance representational capacity, multi-tree GP representation has been explored to evolve more information within a single individual and has been successfully applied to problems such as lung cancer diagnosis [8]. However, most existing GP-based methods focus on a single imaging modality and learn from 2D images extracted from a single anatomical plane. To further improve performance and robustness, ensemble strategies have been incorporated into GP-based image classification. For example, Fan et al. [5] proposed an ensemble GP method that evolves multiple classifiers and uses weighted majority voting for single-modality, single-plane 2D image classification tasks. For 3D image analysis, slice selection has commonly been adopted to reduce computational cost caused by high dimensionality. A recent GP-based study by Wu et al. [9] employed slice selection to learn from 3D inputs, but was also limited to a single plane. These limitations motivate the proposed 3DMPGP framework, which explicitly integrates multi-plane, multi-tree learning with ensemble-based multimodal fusion to comprehensively extract discriminative information from 3D multimodal neuroimaging data.

III. The Proposed Method

A. Proposed Method: 3DMPGP

The overall framework of the proposed 3DMPGP method is shown in Fig. 1. The proposed 3DMPGP method consists of five main components: 1) the 3D MRI and PET images are preprocessed to construct paired inputs and form the multimodal dataset, 2) a multi-tree genetic programming representation is used for multi-plane modeling, where each tree is associated with one anatomical plane, 3) multimodal learning is performed by joint feature learning from 3D MRI and 3D PET data, producing a plane-specific classifier, 4) an ensemble strategy is employed to vote for the predictions from axial, sagittal, and coronal plane, and 5) a frequency-based visualization method is used to analyze the selected brain regions.

B. 3D Data Preprocessing and Preparation

The MRI and PET scans are preprocessed to ensure consistent image quality, spatial alignment, and anatomical correspondence across subjects. For MRI, the preprocessing

TABLE I
THE FUNCTION SET OF 3DMPGP.

Function type	Functions
Classifier	*SVM, RF, LR*
Feature Fusion	*Concat, WAvg, Add*
Feature Concatenation	*FeaCon2v, FeaCon3v, FeaCon4v, FeaCon5v, FeaCon6v*
Feature Extraction	*G_DIF, G_HIST, G_SIFT, G_HOG, G_uLBP, L_DIF, L_HIST, L_SIFT, L_HOG, L_uLBP*
Region Detection	*RegionR, RegionS*
Slice Selection	*Slice_Selection*

pipeline included noise filtering, spatial resampling, bias field correction, and skull stripping. The MRI scans are then aligned and spatially normalized to the MNI152 template with a voxel resolution of 2×2×2 mm, ensuring standardized image dimensions and consistent anatomical locations across subjects. The procedure of preprocessing the PET includes origin set, alignment, skull stripping, spatial normalization, and smoothing. To achieve accurate multimodal correspondence, each PET scan is co-registered to the corresponding preprocessed MRI scan using the MRI as the reference image due to its higher spatial resolution. After completing these preprocessing steps, both MRI and PET scans are processed to a resolution of 91×109x91 within MNI152 space.

Given a preprocessed 3D MRI image, a texture-driven slice indexing strategy is used to identify a set of key 2D slices that have high texture information, which together form a thinner 3D MRI image. The corresponding PET data adopt the same slice indices derived from the MRI to construct the thin 3D PET image, ensuring anatomical correspondence between modalities. The resulting 3D MRI and PET data from the same anatomical plane and spatial location are paired and used as multimodal inputs to a single GP tree. This design maintains spatial consistency across modalities while allowing each anatomical plane to be modeled independently.

C. Multi-tree GP Representation for Multi-plane View

As shown in the **example individual** in Fig. 1, in the proposed 3DMPGP, a multi-tree genetic programming representation is employed to model comprehensive anatomical information from 3D neuroimaging data from three plane views. The 3DMPGP decomposes the 3D image learning task into three plane-specific learning tasks corresponding to the axial, sagittal, and coronal planes, allowing plane-specific discriminative patterns to be learned independently. Each individual in the GP population consists of three trees, with each tree evolving on one anatomical plane. This design explicitly associates each GP tree with a fixed plane, ensuring that spatial characteristics unique to each view are preserved and avoiding interference from other anatomical views during evolution.

a) Multi-tree Crossover and Mutation: In the proposed 3DMPGP method, tournament selection is used to select individuals as parents for generating the next generation. Crossover and mutation are applied separately to the axial, sagittal, and coronal trees. During crossover, a same-index crossover strategy is employed to preserve the structural consistency across planes. Specifically, when two parent individuals are selected for crossover, same-index one-point subtree crossover is applied independently to each of their corresponding trees from the same plane to maintain the learned plane-specific information. The mutation operator replaces a randomly selected subtree with a newly generated subtree, introducing structural variation and promoting diversity in the population.

D. Plane-specific Multimodal Learning and Fusion

As shown in the **example tree** in Fig. 1, within each tree, modality-specific feature learning is performed separately on paired MRI and PET data, followed by multimodal fusion. This design allows the model to jointly learn structural and functional information while preserving their distinct characteristics. The plane-specific tree is complete based on the function set and terminal set.

*a) **Function Set**:* The function set of 3DMPGP is summarized in Table I. For each single 3D image input, the first operation is slice selection, which randomly selects one out of the 9 slices from the 3D image for further feature learning. After a slice is selected, the GP tree determines the region of interest (ROI) to explore - whether select from the entire slice or to focus on a specific region within the slice. Two region selection functions are defined: *RegionR*, which selects a rectangular region, and *RegionS*, which selects a square region.

For feature extraction for each modality, 3DMPGP uses five image descriptors: DIF [10], HIST [11], SIFT [12], HOG [5], and uLBP [13]. The DIF descriptor computes the mean and standard deviation as statistical intensity features, resulting in 20 features in this study. The HIST descriptor extracts features from the image or selected regions by computing normalized intensity histograms. The SIFT descriptor extracts 128-dimensional descriptors based on gradient magnitude and orientation, providing shape-related information. Similarly, HOG captures shape information by computing local gradient distribution patterns. The uLBP descriptor encodes local micro-patterns by thresholding neighborhood pixel intensities against the center pixel and preserves uniform binary patterns, resulting in 59 uniform LBP features in this study, providing texture information. The prefix *G* denotes functions used for global feature extraction, whereas *L* indicates local feature extraction. By combining global information with local details from specific brain regions, 3DMPGP can capture comprehensive characteristics of neuroimaging data to enable effective feature learning.

The last step for the single modality subtree is feature concatenation functions, which are designed to concatenate feature vectors produced by different internal feature extraction nodes for a single modality within a GP tree. In 3DMPGP, a set of feature concatenation functions is defined, which accept two to six feature vectors extracted by different feature extraction operators as inputs.

For multimodal fusion, three feature fusion operators are defined: *Concat*, *WAvg*, and *Add*. The *Concat* operator performs

direct concatenation of MRI and PET feature vectors along the feature dimension, thereby preserving modality-specific information from both sources. The *WAvg* operator uses a learnable weight parameter to control the relative contributions of MRI and PET features, enabling adaptive balancing of structural and functional information. The *Add* operator performs the addition of MRI and PET feature vectors. These operators take the outputs of the MRI and PET subtrees as inputs and produce a fused multimodal feature set for classification.

The last operator to evolve in each tree of 3DMPGP is the classifier. Classifier functions are defined to learn from the fused multimodal feature set to make the predictions. In 3DMPGP, three classifier operators are provided: *SVM*, *LR*, and *RF*, corresponding to Support Vector Machine, Logistic Regression, and Random Forest classifiers, respectively. Each classifier operator takes the fused feature set as input, along with the corresponding labels and a set of learnable hyperparameter, and outputs predicted class labels for that plane.

b) ***Terminal Set****:* For slice selection, an integer-valued terminal specifies the slice index along a given anatomical plane with values ranging from 0 to 8, enabling the extraction of one slice from the input 3D images. For region-based feature extraction, integer-valued terminals are used to define the spatial parameters of the selected regions. Specifically, the terminals *X* and *Y* specify the starting coordinates of a selected region within a slice, with their values constrained to the range [5, 90] to ensure that regions are located within anatomically meaningful areas and to reduce the influence of background regions. The *Length* and *Width* determine the side length of the region, with values ranging from 20 to 50. The only tunable parameter in the multimodal fusion stage is the fusion weight used in the weighted fusion operator, which is sampled uniformly from the interval $[0, 1]$. For the SVM and LR classifiers, an integer-valued terminal is used to control the regularization parameter, with values sampled from the range $[-2, 5]$. For the RF classifier, two integer-valued terminals are employed to specify the number of trees and the maximum tree depth, respectively. The number of trees is sampled from the range $[50, 500]$ with a step size of 10, while the maximum tree depth is sampled from the range $[10, 100]$ with a step size of 10.

E. Ensemble Decision-making Evaluation

Since each anatomical plane captures different patterns, relying on a single plane may lead to incomplete or biased decisions. The ensemble strategy enables the integration of complementary anatomical information from the models across planes while reducing sensitivity to plane-specific errors. To obtain the diagnosis for each subject, the proposed 3DMPGP integrates the predictions from 3 plane-specific classifiers through the multi-plane voting strategy.

As for evaluation, the axial, sagittal, and coronal classifiers from each GP tree make predictions based on the learned multimodal feature set of their respective planes. These predictions are then aggregated using a majority voting strategy to determine the final class label. Compared with single-view or single-tree GP models, this ensemble design provides a more stable and comprehensive understanding of the 3D data from three dimensions for AD diagnosis.

TABLE II
PARAMETER SETTINGS OF THE 3DMPGP METHOD.

Parameter	Value	Parameter	Value
Generations	50	Crossover Rate	0.80
Population Size	100	Mutation Rate	0.19
Initial Population	Ramped Half-and-half	Elitism	0.01
Initial Minimum Depth	5	Tournament Size	5
Initial Maximum Depth	6	Max Tree Depth	9

F. Frequency-based Visualization Analysis

To enhance the interpretability of the proposed 3DMPGP, a frequency-based analysis is used to visualize the brain regions selected by the evolved GP programs across different anatomical planes. First, the GP individual with the highest fitness in each evolutionary run is considered the best-performing individual. Then, the frequency with which brain regions are selected by these top-performing individuals is computed separately for each anatomical plane. After the frequency is normalized, we use a heatmap to visualize. Regions that are repeatedly selected across different individuals and across multiple evolutionary runs are considered to be more discriminative for AD classification. This frequency-based analysis enables the identification of anatomically consistent patterns, revealing how different anatomical views contribute to the final diagnostic decision.

IV. EXPERIMENT DESIGN

A. Dataset

In this study, we have screened a dataset from the Australian Imaging, and Biomarker and Lifestyle Flagship Study of Aging (AIBL) [14] for the training and test process. The number of the subjects is 275, which involved 216 CN subjects and 59 AD patients. Each subject included has both T1-weighted MRI and PiB-PET scans.

Following the preprocessing stage in the proposed method, the processed 3D images are prepared for further analysis. As a result, each image has a size of $9 \times 109 \times 91$ with a voxel size of $2 \times 2 \times 2$ mm. For each subject, the dataset consists of three 3D images extracted along the three anatomical planes from MRI and three corresponding images from PET. All pixels are normalized to the range [0, 1].

B. GP Settings and Evaluations

The parameter settings of the proposed 3DMPGP method are listed in Table II. The evolutionary process continues until either the maximum limit of 50 generations is reached or an individual achieves 100% balanced classification accuracy.

The proposed 3DMPGP method is executed 30 independent runs with different random seeds, and the average values are reported. Given the relatively small number of subjects in the AIBL dataset, we use five-fold cross-validation for overall performance evaluation in each run. During the training process, individuals are evaluated based on their classification performance, and balanced accuracy is used as training fitness.

TABLE III
PERFORMANCE COMPARISON OF DIFFERENT METHODS.

Methods	ACC	BACC	SEN	SPE
FELGP [15]	83.66±1.24 (+)	68.40±6.17 (+)	42.98±4.90 (+)	93.83±2.82 (≈)
3DResNet [16]	82.18±4.66(+)	85.15±2.69(+)	79.884±7.38(+)	90.418±7.82(+)
MDLNET [6]	87.27±2.57 (+)	84.54±8.94 (≈)	79.53±17.70 (≈)	89.55±2.32 (+)
Proposed$_{(MRI)}$	82.03±1.83 (+)	74.18±3.06 (+)	60.42±6.57 (+)	87.95±2.31 (+)
Proposed$_{(PET)}$	88.80±0.92 (+)	83.80±2.19 (+)	74.98±4.86 (+)	92.61±1.08 (≈)
Proposed$_{(Axial)}$	88.15±1.42 (+)	85.45±2.36 (≈)	80.68±5.87 (≈)	90.22±2.34 (+)
Proposed$_{(Sagittal)}$	89.08±1.33 (+)	87.54±2.43 (≈)	84.85±5.47 (–)	90.23±1.79 (+)
Proposed$_{(Coronal)}$	88.71±1.18 (+)	86.22±2.83 (≈)	81.81±6.32 (≈)	90.63±1.37 (+)
Proposed	**89.94±1.36**	**86.74±2.68**	**81.10±5.51**	**92.38±1.27**

For each fold, the best individual with the highest training fitness is selected and used to classify the test data.

To validate the effectiveness of 3DMPGP, it is compared with FELGP, a widely used GP-based ensemble image classification method [15]. FELGP takes raw images as input for feature learning and ensemble decision-making using a majority voting strategy. In this study, FELGP is implemented using the same parameter settings as 3DMPGP to ensure a fair comparison. In addition, 3D ResNet-18 [16] (3DResNet) is a widely used deep learning architecture for volumetric medical image analysis and serves as a strong baseline for end-to-end medical image analysis, incorporating three-dimensional convolutions for 3D image classification. Last, MDLNET [6] is a recent neural network–based method specifically designed for multimodal neuroimaging analysis and has demonstrated good performance in AD diagnosis.

V. RESULTS AND DISCUSSIONS

A. Overall Results

The results of the experiments are presented in Table III. For clarity, the proposed method is highlighted in bold. The mean and standard deviation results of 3DMPGP and other methods over the 30 runs on the investigated dataset are shown with the test accuracy (ACC), balanced accuracy (BACC), sensitivity (SEN), and specificity (SPE). To assess the statistical significance of performance differences, the Wilcoxon rank-sum test with a 95% significance interval is applied. The symbols "$(+)$" and "$(-)$" in Table III indicate whether our proposed 3DMPGP method achieves significantly better or worse performance than the compared method.

a) Performance Comparison with Baseline Methods: The first section in Table III presents the overall performance of the baseline methods. Compared with the GP-based ensemble method FELGP, the proposed method shows clear improvements on this class imbalance task, particularly in sensitivity, indicating that incorporating multimodal learning and the multi-plane voting strategy substantially enhances 3D image classification performance. Compared with 3D ResNet, 3DMPGP also achieves higher performance overall. When compared with the AD-specific 3D image classification method MDLNET, the proposed method achieves higher test accuracy and specificity. In addition, the proposed method maintains the higher specificity while achieving competitive sensitivity, suggesting improved ability in diagnosing AD.

b) Ablation Study on Single-Modality 3DMPGP: The second section in the table demonstrates that the proposed 3DMPGP method consistently benefits from multimodal learning. Compared with the single-modality 3DMPGP, the proposed multimodal 3DMPGP method achieves the highest overall performance across all evaluation metrics, with statistically significant improvements in test accuracy, balanced accuracy, and sensitivity. This indicates that fusing structural information from MRI with functional information from PET provides more discriminative information than using a single modality. In particular, the MRI 3DMPGP shows significantly lower sensitivity, suggesting difficulty in correctly identifying AD cases when relying only on structural information. Although the balanced accuracy of PET 3DMPGP is statistically similar to that of multimodal 3DMPGP, it still underperforms other evaluation metrics, highlighting that the complementary information of MRI and PET used in 3DMPGP for AD diagnosis is necessary.

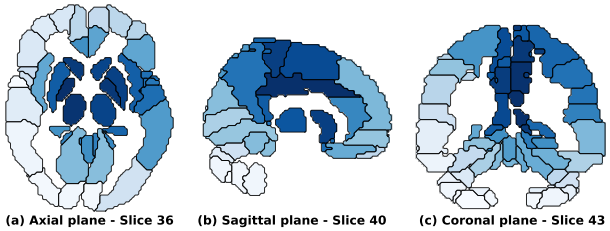

Fig. 2. Region-wise Frequencies in Axial, Sagittal, and Coronal Planes

c) Ablation Study on Single-Plane 3DMPGP: The third section in the table presents the results across different planes. The results show that the proposed 3DMPGP model achieves competitive and stable performance, confirming the effectiveness of modeling 3D data through multiple 2D planes. Among the single-plane models, sagittal-plane learning achieves the highest sensitivity, suggesting that the model learns clearer disease-related patterns along this anatomical dimension. However, no single-plane model outperforms the proposed 3DMPGP method in terms of balanced accuracy and specificity, reflecting the spatial variability of AD-related changes in the brain. These results demonstrate that learning from multiple planes enables a more effective and comprehensive understanding of 3D neuroimaging data than single-plane learning alone.

B. Brain Region Analysis

Fig. 2 visualizes the most frequently selected slices and the spatial distribution of brain regions most commonly identified by the evolved GP trees across the sagittal, coronal, and axial planes, based on the AAL brain atlas [17]. The heatmap shows frequency distributions, where darker colors indicate higher frequencies. Clear plane-dependent patterns can be observed, indicating that different anatomical planes emphasize distinct brain regions during feature learning.

As shown in Table IV, the sagittal and coronal planes have overlap in the regions most frequently selected by the solutions evolved 3DMPGP, particularly within the cingulate cortex. The middle cingulate and paracingulate gyri on the left and right brain are separately selected by two planes, indicating that these planes consistently capture the information that is strongly associated with AD-related cognitive and memory impairments, which have been proven to be consistent with previous AD studies [18], [19]. This cross-plane consistency suggests that sagittal and coronal views provide comprehensive

TABLE IV
TOP THREE MOST FREQUENTLY SELECTED BRAIN REGIONS.

Plane	Rank	AAL Label	Region Name
Sagittal	1	38	Cingulate_Mid_R
Sagittal	2	76	Caudate_R
Sagittal	3	74	Paracentral_Lobule_R
Coronal	1	39	Cingulate_Post_L
Coronal	2	37	Cingulate_Mid_L
Coronal	3	71	Precuneus_L
Axial	1	80	Pallidum_R
Axial	2	82	Thalamus_R
Axial	3	79	Pallidum_L

perspectives on similar disease-relevant regions, reinforcing their importance during the evolving process. In contrast, the axial plane highlights a different set of regions, with the pallidum and thalamus being most frequently selected, which is also consistent with previous studies [20], [21]. This difference indicates that axial-plane learning contributes additional information related to the brain structures that may not be effectively captured from other planes. Together, these findings demonstrate the necessity of multi-plane learning to capture cross-plane and cross-structure relationships in 3D neuroimaging data.

VI. CONCLUSIONS

This paper proposed an ensemble multimodal multi-plane multi-tree GP method for Alzheimer's disease diagnosis using 3D MRI and PET data. By learning complementary structural and functional information from multimodal inputs and modeling 3D neuroimaging data across axial, sagittal, and coronal planes, the proposed method effectively addresses the challenges posed by high-dimensional volumetric data and the limited number of samples. Plane-specific GP models are evolved independently, and then their predictions are combined through majority voting, allowing comprehensive information from different plane views to be fused at the decision level. Experimental results demonstrate that this multi-plane multimodal ensemble strategy leads to improved robustness and overall classification performance compared with single-modality and single-plane models, as well as competitive baseline methods. The findings in this paper highlight the effectiveness of combining a multi-plane voting strategy with multi-tree GP for 3D multimodal neuroimaging AD diagnosis. Although the proposed method achieves competitive overall performance, its sensitivity is still relatively limited, indicating that further work is needed to improve the identification of AD subjects, particularly under class-imbalanced settings.

REFERENCES

[1] L. Yin, C. Ye, T. Liu, J. Wu, and T. Yan, "Unicross: Balanced multimodal learning for alzheimer's disease diagnosis by uni-modal separation and metadata-guided cross-modal interaction," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2025, pp. 638–648.

[2] Y. Zhang, F. Zhang, B. Xue, and M. Zhang, "Multimodal image classification using genetic programming for alzheimer's disease diagnosis," in *IEEE Congress on Evolutionary Computation*. IEEE, 2025, pp. 1–8.

[3] M. L. Raza, S. T. Hassan, S. Jamil, N. Hyder, K. Batool, S. Walji, and M. K. Abbas, "Advancements in deep learning for early diagnosis of alzheimer's disease using multimodal neuroimaging: challenges and future directions," *Frontiers in Neuroinformatics*, vol. 19, p. 1557177, 2025.

[4] D. Zhang, Y. Wang, L. Zhou, H. Yuan, D. Shen, A. D. N. Initiative *et al.*, "Multimodal classification of alzheimer's disease and mild cognitive impairment," *Neuroimage*, vol. 55, no. 3, pp. 856–867, 2011.

[5] Q. Fan, Y. Bi, B. Xue, and M. Zhang, "A multi-tree genetic programming-based ensemble approach to image classification with limited training data [research frontier]," *IEEE Computational Intelligence Magazine*, vol. 19, no. 4, pp. 47–62, 2024.

[6] Z. Qiu, P. Yang, C. Xiao, S. Wang, X. Xiao, J. Qin, C.-M. Liu, T. Wang, and B. Lei, "3d multimodal fusion network with disease-induced joint learning for early alzheimer's disease diagnosis," *IEEE Transactions on Medical Imaging*, vol. 43, no. 9, pp. 3161–3175, 2024.

[7] S. Mohsen, "Alzheimer's disease detection using deep learning and machine learning: a review," *Artificial Intelligence Review*, vol. 58, no. 9, p. 262, 2025.

[8] E. Scalco, G. Rizzo, and W. Gómez-Flores, "Automatic feature construction based on genetic programming for survival prediction in lung cancer using ct images," in *Annual International Conference of the IEEE Engineering in Medicine & Biology Society*. IEEE, 2022, pp. 3797–3800.

[9] Z. Wu, B. Xue, and M. Zhang, "Multitree genetic programming for multimodal learning in multimodal medical image classification," *IEEE Transactions on Evolutionary Computation*, 2025.

[10] M. Zhang, V. B. Ciesielski, and P. Andreae, "A domain-independent window approach to multiclass object detection using genetic programming," *EURASIP Journal on Advances in Signal Processing*, vol. 2003, pp. 1–19, 2003.

[11] A. Vedaldi and B. Fulkerson, "Vlfeat: An open and portable library of computer vision algorithms," in *Proceedings of the 18th ACM international conference on Multimedia*, 2010, pp. 1469–1472.

[12] N. Zayed, G. Eldeep, and I. A. Yassine, "Classification method based on surf and sift features for alzheimer diagnosis using diffusion tensor magnetic resonance imaging," *Scientific Reports*, vol. 15, no. 1, p. 9782, 2025.

[13] K. Prakash and S. Saradha, "Efficient prediction and classification for cirrhosis disease using lbp, glcm and svm from mri images," *Materials Today: Proceedings*, vol. 81, pp. 383–388, 2023.

[14] K. A. Ellis, A. I. Bush, D. Darby, D. De Fazio, J. Foster, P. Hudson, N. T. Lautenschlager, N. Lenzo, R. N. Martins, P. Maruff *et al.*, "The australian imaging, biomarkers and lifestyle (aibl) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of alzheimer's disease," *International psychogeriatrics*, vol. 21, no. 4, pp. 672–687, 2009.

[15] Y. Bi, B. Xue, and M. Zhang, *Genetic programming for image classification: An automated approach to feature learning*. Springer Nature, 2021.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[17] E. T. Rolls, M. Joliot, and N. Tzourio-Mazoyer, "Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas," *Neuroimage*, vol. 122, pp. 1–5, 2015.

[18] S. Pezzoli, J. Giorgio, X. Chen, T. J. Ward, T. M. Harrison, and W. J. Jagust, "Cognitive aging outcomes are related to both tau pathology and maintenance of cingulate cortex structure," *Alzheimer's & Dementia*, vol. 21, no. 2, p. e14515, 2025.

[19] Q. Zhao, X. Sang, H. Metmer, J. Lu, A. D. N. Initiative *et al.*, "Functional segregation of executive control network and frontoparietal network in alzheimer's disease," *Cortex*, vol. 120, pp. 36–48, 2019.

[20] L. W. de Jong, K. van der Hiele, I. M. Veer, J. Houwing, R. Westendorp, E. Bollen, P. W. de Bruin, H. Middelkoop, M. A. van Buchem, and J. van der Grond, "Strongly reduced volumes of putamen and thalamus in alzheimer's disease: an mri study," *Brain*, vol. 131, no. 12, pp. 3277–3285, 2008.

[21] T. Iizuka, Y. U. Kameyama, M. Fukasawa, M. Akishita, and M. Kameyama, "Impaired ability to smile in alzheimer's disease is associated with reduced brain volume of nucleus accumbens and pallidum, revealed by artificial intelligence," *Alzheimer's & Dementia*, vol. 19, p. e082571, 2023.