Multimodal Image Classification Using 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

2025 IEEE Congress on Evolutionary Computation (CEC) | 979-8-3315-3431-8/25/$31.00 ©2025 IEEE | DOI: 10.1109/CEC65147.2025.11042959

***Abstract*—Alzheimer's disease (AD) is a progressive neurological disorder and a major contributor to dementia cases across the world. Timely and accurate diagnosis is crucial for effective clinical management and therapeutic intervention. This paper presents a genetic programming (GP) method with a multi-tree representation designed to effectively integrate multimodal neuroimaging data while preserving spatial information for AD classification. Unlike existing GP approaches that focus on single-modality data, our GP approach directly uses the images from multiple imaging sources as inputs into the evolutionary process. A new GP representation is designed to handle multimodal data effectively, enabling feature extraction and classification. Experiments on the commonly used public database of Alzheimer's disease neuroimaging initiative (ADNI) show that the proposed method performs effectively in diagnosing AD. These findings suggest that multi-tree GP has the potential to serve as a powerful and interpretable tool for neuroimaging-based AD diagnosis, offering a promising approach to improve AD detection and clinical decision-making.**

***Index Terms*—Genetic programming, image classification, neuroimaging, multimodal classification, medical diagnosis**

I. Introduction

Alzheimer's disease (AD) is one of the most common progressive neurodegenerative disorders, gradually impairing cognitive functions and ultimately leading to severe dementia and death. In 2006, the worldwide number of Alzheimer's disease cases was estimated at 26.6 million, and this figure is expected to reach 106.8 million by 2050 due to the growing aging population [1]. Thus, accurate early detection of AD is crucial for improving patient outcomes through timely clinical intervention and effective monitoring of disease progression. However, AD presents a prolonged asymptomatic preclinical phase, during which individuals who appear cognitively normal may still have the disease [2]. This complexity makes the diagnosis of AD a challenging task. Despite the difficulties, findings from neuroimaging studies have provided valuable insights, leading to pathological diagnoses that identify certain brain features as highly indicative of AD, with moderate to severe cortical atrophy often observed [3].

Neuroimaging techniques, including magnetic resonance imaging (MRI) and positron emission tomography (PET) [4], are commonly used for evaluating and diagnosing AD, each offering unique advantages but with limitations. MRI is commonly used due to its high resolution, which allows for a clear display of soft tissues in the brain, like the subtle structures and abnormalities. However, it cannot provide functional information related to metabolic activity. On the other hand, PET is an essential imaging modality that provides valuable insights into the metabolic processes associated with AD. Despite its ability to capture functional changes, PET suffers from lower resolution compared to MRI, which can limit the precision of details. In general, the MRI and PET provide complementary advantages in AD diagnosis.

With multiple data modalities, the AD classification still faces several critical hurdles. One of the primary issues is the limited availability of data. In terms of data availability, the Alzheimer's Disease Neuroimaging Initiative (ADNI) database [5], one of the most widely used datasets for AD research, provides only approximately 300 patients who have both MRI and PET imaging data. Besides, MRI and PET data are all 3D images, consisting of millions of pixels, which significantly increases computational complexity and cost, and poses challenges for classification algorithms.

Traditional machine learning classification follows a two-stage process: feature extraction from images, followed by a classification algorithm using extracted features to make predictions. Traditional machine learning algorithms, such as the Support Vector Machine (SVM) [6], are widely employed due to their effectiveness in handling multimodal medical tasks. However, these classification algorithms face challenges related to the curse of dimensionality and, as a result, they normally do not directly utilize raw 3D images as input. Instead, they rely on a predefined, fixed number of features extracted from specific brain regions using one domain-specific technique, such as volume measurements for MRI and voxel intensity values for PET [6]. Even deep learning models, which rely on large training datasets to learn patterns, face significant challenges, like computational costs, when handling high-dimensional 3D medical imaging data. Thus, slice selection is a commonly used method to improve efficiency [7]. Although deep neural networks demonstrate strong performance in classification tasks, the low transparency due to their black-box nature makes it challenging to interpret the decision-making process, raising concerns about their reliability, particularly in the medical field [8].

Genetic Programming (GP) is an evolutionary algorithm that can automatically evolve image descriptors with a flexi-

ble variable length representation [9]. This flexibility allows GP to explore complex solution spaces more effectively by dynamically adjusting the evolved programs' structure and complexity to fit different data sources' characteristics better. Besides, the tree-based representation of GP offers a high level of understandability [10]. Compared with deep learning methods, GP demonstrates a strong capability to learn from small datasets, and the solution of GP is much easier to analyze and comprehend the underlying decision-making process, making it particularly suitable for medical applications [11]. Different from standard GP, which evolves a single tree in an individual, multi-tree GP uses multiple trees within an individual to capture various aspects of information to solve the Melanoma classification problem [12]. This approach allows for a more comprehensive representation by enabling each tree to focus on distinct features or modalities, making it particularly well-suited for tasks that require the fusion of multiple data sources. Moreover, the GP algorithm can evolve both end-to-end models [13] or take advantage of traditional machine learning classifiers [10], which combine the strengths of evolutionary search with established machine learning techniques, leading to improved performance.

While GP has demonstrated its potential in single-modality image classification, the current literature using GP to solve the problem of multimodal medical image classification remains relatively unexplored. The complexity of integrating multiple imaging modalities and the need for high accuracy and transparency present unique challenges that require further investigation. Therefore, it is essential to explore the potential of GP on this task and develop GP-based approaches that can enhance classification performance and provide greater transparency to support AD diagnosis.

This paper aims to develop a multi-tree GP-based method for AD classification using multimodal neuroimaging data, mainly MRI and PET. Unlike current GP methods for processing multimodal data [14], which use pre-extracted features as input, our method directly works on raw images. Additionally, the proposed method effectively integrates complementary structural and functional information by employing a multi-tree representation, where different trees process specific brain slices for one modality within a single individual. This design ensures that crucial modality-specific patterns are preserved. Furthermore, the method captures critical information from both the entire image and small regions by automatically evolving feature extraction within GP, rather than relying on predefined handcrafted feature engineering. By fusing the extracted features from each modality, the proposed approach aims to utilize complementary information from both modalities to improve classification accuracy and generalization performance for AD diagnosis. The overall goals consist of the following specific objectives:

- Propose a new GP method to directly and simultaneously evolve multimodal representations for AD classification by effectively integrating MRI and PET data.
- Evaluate the effectiveness of multimodal data by comparing the performance of multiple modalities against a single modality to demonstrate the advantages of multimodal images.
- Identify important feature descriptors that can help AD classification by capturing relevant characteristics from regions or whole images within the GP program structure.

The remainder of the paper is organized as follows. Section II discusses the background and related work, while Section III describes the proposed new methodology. In Section IV, the experiment settings are discussed. The results of the proposed GP algorithm on the AD datasets are presented in Section V. Finally, Section VI concludes the paper and highlights some future directions.

II. Background and Related Work

A. AD Diagnosis and Multimodal Image Classification

AD is a progressive neurodegenerative disorder that gradually impairs cognitive function. To assess the severity of cognitive decline, clinicians commonly rely on various cognitive tests, such as the Clinical Dementia Rating (CDR) and the Mini-Mental State Examination (MMSE). However, brain changes may occur before the first clinical signs of AD appear. As a result, neuroimaging has emerged as a highly promising tool for early diagnosis and monitoring AD.

Multimodal image classification, as defined in this paper, involves classifying different classes, such as cognitively normal (CN) and AD subjects, by integrating multiple types of imaging data, such as MRI and PET, to enhance diagnostic accuracy. Single-modality neuroimaging data typically provide only partial information about brain abnormalities, which may theoretically limit classification performance.

To overcome this limitation, several machine learning methods have been proposed using multimodal neuroimaging to improve classification accuracy. Zhang et al. [6] used an SVM classifier combined with volumetric features extracted from MRI and PET, demonstrating improved performance compared to single-modality models. However, approaches of this kind rely on pre-extracted features rather than directly using raw images, which may limit their flexibility in capturing complex and diverse patterns in multimodal imaging data. Deep learning methods that use images as input have also been widely applied in multimodal neuroimaging. Liu et al. [15] proposed a convolutional neural network architecture that automatically learns features from MRI and PET images, outperforming traditional hand-crafted feature-based methods. However, most of these methods are prone to overfitting, exhibiting high training accuracy but significantly lower accuracy on the testing set due to limited data availability.

B. GP for Image Classification

Genetic Programming (GP) has achieved notable success in image classification due to its ability to automatically evolve feature representations. Bi et al. [16] proposed a GP-based algorithm named FLGP to evolve solutions using global feature extraction methods and local feature extraction methods to extract features for image classification. This approach uses images as input and demonstrated improved performance

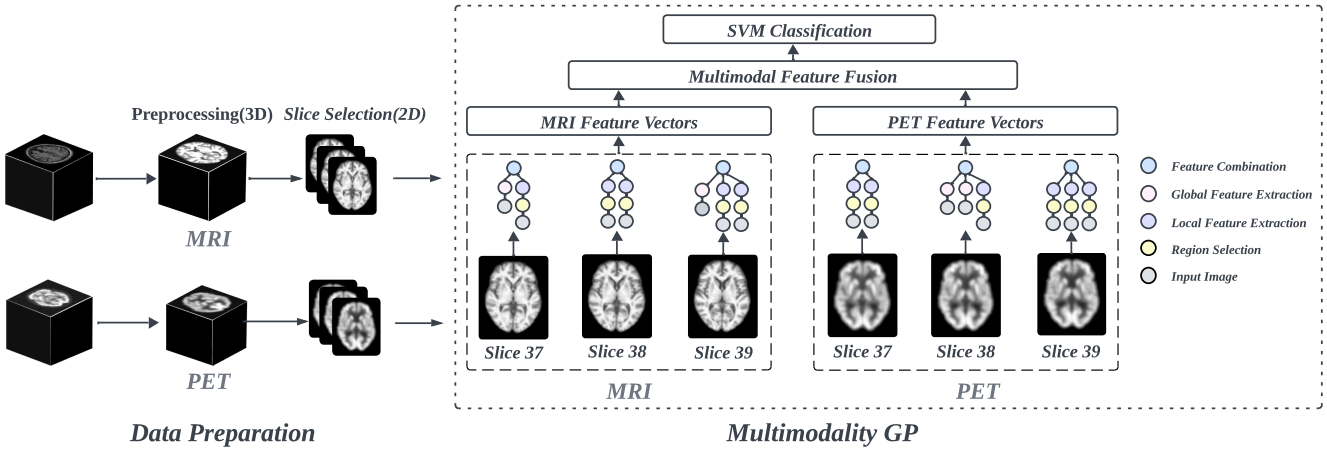

Fig. 1. Overview of the proposed MMTGP approach.

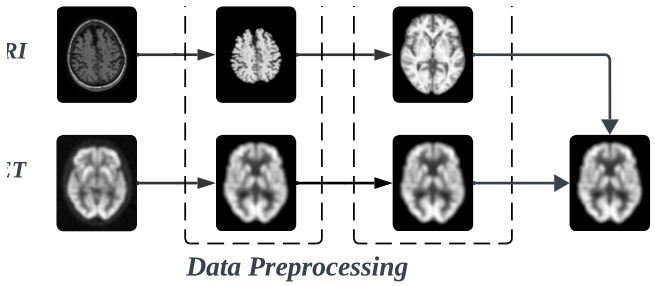

Fig. 2. The detail of preprocessing steps using one slice as an example.

compared to methods relying solely on hand-crafted features, demonstrating GP's potential to adaptively learn discriminative representations. However, FLGP is limited to single-modality data and lacks the mechanisms required for processing and fusing different data modalities, making it less effective for multimodal classification tasks theoretically.

Extending GP from single-modality image classification to multimodal classification presents several challenges. First, the heterogeneity of data sources requires the GP model to effectively process and handle distinct information from each modality. Additionally, successful multimodal classification involves preserving the unique characteristics of each modality during the fusion process to fully use their complementary strengths. However, existing GP methods struggle to achieve that. For example, Ain et al. [14] designed a two-stage GP approach for multimodal datasets to diagnose skin cancers. While this method integrates multiple modalities, it relies on pre-extracted LBP image descriptors as input to the GP model, limiting its flexibility to operate directly on raw image data and to discover complex, modality-specific feature representations automatically.

For AD classification, Tang et al. [17] developed a GP-based method using data from the ADNI database. However, this method processes only a single modality and uses pre-extracted features as input to perform the regression task rather than directly integrating multimodal information for classification.

To summarize, existing GP-based approaches have demonstrated success in single-modality image classification tasks but face limitations in fully utilizing raw multimodal imaging data and efficiently extracting complementary features across modalities for AD classification. To address these challenges, this study proposes a new GP-based method designed specifically for multimodal AD image classification.

III. The Proposed Method

This section describes the proposed Multimodal Multi-Tree Genetic Programming (MMTGP) method, whose structure is illustrated in Fig. 1. The MMTGP framework consists of two main components: data preparation and multimodality GP evolution. In the data preparation phase, the raw 3D neuroimaging data undergoes preprocessing to convert it into standardized 3D image data, removing irrelevant information. To reduce computational complexity, we perform slice selection to extract key 2D slices from the 3D data. These selected slices represent the same regions of the brain across modalities, ensuring consistency when integrating information from the two sources. In the multimodality GP evolution phase, each tree in the MMTGP structure evolves from a single slice of the corresponding modality. With an input image, a GP tree can generate a set of features by employing global and/or local feature extraction operators to either the whole image or specific regions identified by the region selection function. These extracted features are then combined into a single feature vector as output. The outputs of the trees from each modality are then combined to perform multimodal feature fusion. The fused feature representations are subsequently passed to a SVM classifier for final classification. This integrated process ensures effective feature extraction, fusion, and classification for accurate Alzheimer's disease diagnosis.

A. Data Preprocessing and Slice Selection

MRI and PET images provide complementary information that is essential for detecting AD and its various stages. MRI primarily captures structural details, while PET provides functional insights, making both modalities critical for

comprehensive automated diagnosis. However, differences in acquisition parameters, slice counts, and slice sizes between MRI and PET scans necessitate standardized preprocessing to ensure the integration.

In this study, the MRI and PET data from the ADNI database are preprocessed using the Statistical Parametric Mapping (SPM12) toolbox [18], which are shown in Fig. 2. The skull stripping is performed on all 3D MRI and PET scans to remove the skull and other non-brain tissues, resulting in brain-only scans with reduced noise and irrelevant information. Following this, the scans are affine-transformed to the MNI152 space [19], a universal brain atlas template, to eliminate spatial discrepancies between subjects, minimizing variations in orientation, translations, and rotations. Co-registration is performed using the MRI scans as a reference to align the two modalities accurately due to the high resolution of MRI images. During co-registration, each PET scan is aligned slice by slice to match the MRI reference, ensuring consistent spatial orientation, image size, and voxel dimensions for accurate anatomical correspondence between the two modalities. After these preprocessing steps, both MRI and PET scans are standardized to the same size, $91 \times 109 \times 91$.

Processing the entire 3D brain scan is computationally intensive and time-consuming, making it hard for AD classification. To address these challenges, key slice selection is used as an efficient method. Entropy values, computed using the gray level co-occurrence matrix, are employed to extract texture features that represent the most informative regions of the image, particularly slices impacted by atrophy. By counting the occurrences of the slice with the highest entropy, we identified and included three slice indices, i.e., Slice 37, 38, and 39. Based on this, three significant slices(index 37, 38, 39) from MRI, and three from PET (index 37, 38, 39) for each subject are selected. In total, our dataset has 840 2D images from 140 subjects, which are then used as input for GP.

B. Program Representation and Evaluation

This study employs a multi-tree GP representation to address the complexity of multimodal data. This representation's overall structure and workings are illustrated in Fig. 1, providing a clear visual explanation of the multi-tree approach. In this approach, each individual in the GP population consists of six trees: three trees for *MRI* and the remaining three trees for *PET*. Using different trees for each modality allows the model to capture and maintain modality-specific features and characteristics. Specifically, six trees are employed, with three dedicated to extracting features from slices 37, 38, and 39 of the MRI and three for the corresponding slices of the PET

When doing the multimodal feature fusion, there are three steps. Firstly, the output feature vectors from all three trees assigned to a single modality are concatenated to form a comprehensive representation of the information learned from that modality. Then, normalization is performed to the feature vectors from each modality to ensure that features from different modalities are on a comparable scale. This operation is performed separately for each modality before concatenation to preserve modality-specific patterns. After that, we concatenate the two normalized modality features into one feature vector. This feature vector contains information from both modalities and is then fed into a linear SVM to do classification. Given the relatively small and balanced dataset with a roughly similar number of subjects in each class, the cross-validation classification accuracy on the training set is used as training fitness to fully use the limited data.

TABLE I
FUNCTION SET OF MMTGP

Function type	**Functions**
Region Detection	*RegionR, RegionS*
Feature Extraction	*G_DIF, G_HIST, G_SIFT, G_HOG, G_uLBP, L_DIF, L_HIST, L_SIFT, L_HOG, L_uLBP*
Feature Combination	*FeaCon2v, FeaCon3v*

During the test phase, the best individual with the highest fitness during training transforms the images from both the training and test sets into feature vectors. These feature vectors are then normalized using the min-max normalization method, in which the test set normalization is performed based on the parameters derived from the training set. Subsequently, the normalized training set is used to train a linear SVM classifier, which is then evaluated on the normalized test set. The test data remains independent of the training process, and the entire process is repeated 5 times, which is the 5-fold cross-validation. The average results across the five folds are reported as the performance of the algorithms.

C. Function Set and Terminal Set

The function set of MMTGP includes region selection functions, feature extraction functions, and feature combination functions, as shown in Table I. For region selection, two key functions are used: *RegionR* and *RegionS*, which are designed to select rectangular or square regions from the input image. Both functions take five parameters as input: *Image*, *X*, and *Y*, *Length*, *Width*. The input image is denoted by *Image*, while *X* and *Y* specify the coordinates of upper-left corner of the selected region. Once the starting point is determined, the area of a rectangle is specified by *Length* and *Width*, representing the width and height of the rectangular region. When using *RegionS* to select a square region, the values of *Length* are used as the length and width, which are equal for a square region.

For feature extraction, the MMTGP uses five feature extraction methods, enabling it to extract diverse features. The five methods are DIF [20], HIST [21], SIFT [22], HOG [23], and uLBP [12]. For global feature extraction, which operates on the whole image, the following feature extraction functions are used: *G_DIF*, *G_HIST*, *G_SIFT*, *G_HOG*, and *G_uLBP*. For the *G_DIF*, 20 features are extracted by computing the mean and standard deviation of pixel intensities. The *G_HIST* function computes intensity histograms, providing statistical information about pixel intensity distributions across the image. The *G_SIFT* function extracts 128 SIFT features based on gradient magnitude and orientation, providing a robust representation of shape-related characteristics in brain images.

The *G_HOG* function extracts features that describe shapes by analyzing gradient distributions. The *G_uLBP* function extracts 59 uniform LBP features, which capture texture information—an essential property for brain images, particularly in MRI. For local feature extraction, which focuses on selected regions of interest, five functions are used: *L_DIF*, *L_HIST*, *L_SIFT*, *L_HOG*, and *L_uLBP*. These local feature extraction functions operate similarly to the global functions but are restricted to selected regions within the image as input, allowing MMTGP to evolve more detailed representations that capture both global information and localized information on MRI and PET images. Thus, the MMTGP structure effectively captures information within each modality, enhancing MMTGP's ability to use multimodal information to do AD classification.

Feature combination functions aim to concatenate feature vectors extracted by various internal feature extraction nodes of GP trees. Within the MMTGP, two functions are utilized: *FeaCon2v* and *FeaCon3v*, which accept two or three feature vectors as inputs, respectively. It is worth mentioning that a feature combination function can function as a child node beneath another combination function, which operates as the root node of the tree. This flexibility allows GP trees to dynamically construct complex feature representations. As a result, MMTGP can effectively adapt to varying levels of modality complexity by utilizing trees with flexible depths, allowing the construction of feature representations with varying numbers of features. This flexibility ensures that the model can capture both simple and complex patterns, making it well-suited for representing different data modalities. As mentioned earlier, the parameters required for feature extraction methods must be properly defined. In the MMTGP approach, the *Terminals* include the parameters *Image*, *X*, *Y*, *Length*, and *Width*, which are essential for specifying the input image and the selected region from which features are extracted. The *Image* parameter represents a two-dimensional grayscale MRI or PET image within each tree. The *X* and *Y* parameters define the starting point of the selected region, with their values constrained to the range [0, width of the input image − 20] and [0, height of the input image − 20], respectively. This constraint ensures that the extracted region remains within the image boundaries while also reducing the influence of black areas that do not contain brain tissue. The *Length* and *Width* parameters determine the size of the selected regions, with their values ranging from 20 to 50, allowing the extracted region to vary between 20 × 20 and 50 × 50 pixels. This flexibility ensures that the MMTGP can adaptively extract meaningful local features from different regions of the image, enhancing the effectiveness of multimodal feature representation for classification. If *X* + *Length* exceeds the image width, or if *Y* + *Width* exceeds the image height, the region is automatically clipped to the image boundaries.

D. Crossover and Mutation

In the proposed MMTGP, individuals are selected as parents for the new population generation using the tournament selection method, which chooses the individual with the highest fitness out of the randomly selected individual subset. Then, the crossover and mutation operations are applied to evolve individuals over generations. The same-index crossover operation is performed between pairs of parent individuals, exchanging subtrees across the six trees to generate new offspring. Specifically, for each pair of individuals selected for crossover, the operator is applied separately to each of their six trees and has the same index to maintain consistency within each modality within the same brain slice. Besides, the mutation is applied to each tree with a predefined probability. The mutation operator modifies each tree by introducing random changes, enabling diversity in multiple modalities. Similar to crossover, the mutation is applied independently to all six trees.

IV. Experiment Design

A. Dataset

We use the ADNI database [5] for our experimentation. The dataset used in this study includes both males and females with a follow-up during the last 18 months with an age range of 55 to 90 years. In this paper, ADNI subjects with both MRI and PET baseline data are included. Some subjects are removed for some reason, like low quality. Our dataset has a total of 140 subjects, including 67 AD patients and 73 cognitively normal subjects.

The MRI and PET images in the ADNI database have already undergone several standardized preprocessing steps to ensure consistency and reliability across different imaging sites and equipment. Specifically, MRI images have been processed through a series of corrections by ADNI, including Gradwarp, B1 non-uniformity correction, and N3 bias field correction. Similarly, PET images have been subjected to preprocessing steps such as co-registration, averaging, standardization, and uniform resolution adjustments. Further details regarding these procedures can be found on the ADNI website (https://adni.loni.usc.edu/data-samples/adni-data/neuroimaging/). These preprocessing corrections vary depending on the imaging manufacturer and the specific system's RF coil configurations.

Following the preprocessing stage in the proposed method, the acquired 3D images are prepared for further analysis. As a result, each image has a spatial resolution of 91 × 109 × 91 with a voxel size of 2 mm × 2 mm × 2 mm. Given the high dimensionality of 3D medical images, a slice selection method is employed to extract the most informative slices, thereby reducing computational complexity while retaining critical diagnostic information. The dataset consists of three selected slices from MRI and three from PET for each subject, resulting in 840 images. The selected images have a resolution of 109 x 91 pixels, and all pixel values are normalized to the range [0, 1].

There are three extracted datasets used in this study: $Dataset_{MRI}$, $Dataset_{PET}$, and $Dataset_{MRI\_PET}$. The number of images of these datasets is summarized in Table II. Each dataset contains data from 140 subjects, with the same 112

TABLE II
SUMMARY OF THE EXTRACTED DATASETS FOR TRAINING AND TEST

Dataset	Total Images	Training Set	Test Set
$Dataset_{MRI}$	420	336	84
$Dataset_{PET}$	420	336	84
$Dataset_{MRI\_PET}$	840	672	168

TABLE III
PARAMETER SETTINGS OF THE GP 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
Tree Minimum Depth	2	Tournament Size	7
Tree Maximum Depth	4	Max Depth	7

subjects used for training and the same 28 subjects for testing. For example, when the random seed is fixed, if a subject has three MRI slices included in the training set of $Dataset_{MRI}$, then the corresponding three PET slices for the same subject are included in the training set of $Dataset_{PET}$. Furthermore, all six slices (three from MRI and three from PET) for that subject are included in the training set of $Dataset_{MRI\_PET}$. This consistent subject-wise partitioning ensures proper alignment and comparability across the multimodal datasets.

B. GP Settings

The parameter settings of the proposed multi-tree GP method are listed in Table III. The evolutionary process continues iterating until a predefined termination criterion is satisfied. The process concludes either when the maximum limit of 50 generations is reached or when an individual achieves 100% classification accuracy.

In the experiments, 5-fold cross-validation is used for evaluation. The evolved GP individual consists of six trees, and the best individual with the highest training accuracy is used to classify the test data. This process is repeated five times in a single run, and the average training fitness and test accuracy are recorded. Each GP method is executed 30 times with different random seeds, producing 30 training fitness and test accuracy values for analysis.

To validate the effectiveness of MMTGP, we compare it with FLGP, a state-of-the-art GP-based image classification method [23]. FLGP employs a single-tree representation for automatic feature engineering, extracting features that are then classified using SVM. In this study, FLGP is implemented with the same function set, terminal set, parameter settings, and evaluation process as MMTGP to ensure a fair comparison.

V. RESULTS AND DISCUSSIONS

A. Overall Results

The results of the experiments are presented in Table IV. For clarity, the best-performing method for each evaluation metric is highlighted in bold. The classification results of MMTGP and other methods over the 30 runs on the investigated dataset are shown with the maximum accuracy, and the mean and standard deviation. $FLGP_{MRI}$ refers to the FLGP model trained on $Dataset_{MRI}$, with similar naming conventions applied to other models trained on their respective datasets. Specifically, the MMTGP is trained and evaluated on the $Dataset_{MRI\_PET}$. 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 IV indicate whether MMTGP achieves significantly better or worse performance than the compared method, specifically, a "↑" denotes that MMTGP outperforms the compared method with statistical significance. Besides, the convergence curves using average training performance over the 30 runs for all methods are shown in Fig. 3.

TABLE IV
PERFORMANCE COMPARISON OF DIFFERENT METHODS.

	Training		Test	
Method	**Max**	**Mean±Std**	**Max**	**Mean±Std**
$FLGP_{MRI}$	81.07	78.75±0.92 (↓)	66.68	61.24±2.73 (↓)
$FLGP_{PET}$	88.81	87.18±1.03 (↓)	77.38	72.19±2.94 (↓)
$FLGP_{MRI\_PET}$	79.34	77.59±0.80 (↓)	71.78	66.19±2.22 (↓)
$MTGP_{MRI}$	93.93	91.60±1.31 (↓)	75.00	67.98±3.49 (↓)
$MTGP_{PET}$	**99.29**	**97.73±0.76** (↑)	85.72	76.12±4.32 (↓)
MMTGP	98.75	96.94±0.95	**88.57**	**79.31±4.64**

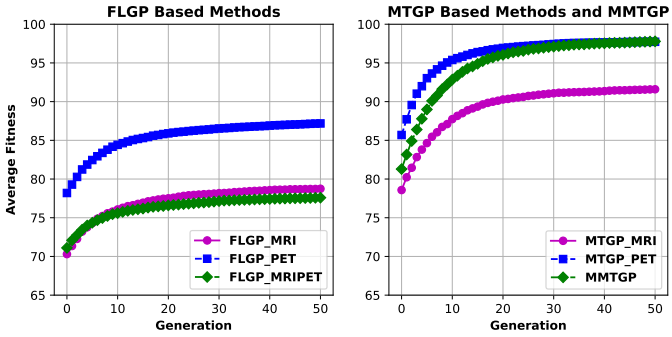

Fig. 3. Convergence curves of FLGP based methods, MTGP based methods, and MMTGP over the 30 runs on three datasets.

The first three rows in Table IV present the classification results of FLGP on the three datasets. For example, the $FLGP_{MRI}$ is trained and tested on $Dataset_{MRI}$. Compared with FLGP based methods, the performance of MMTGP across all datasets is statistically better on both the training and test sets. Within the FLGP-based methods, it is theoretically expected that $FLGP_{MRI\_PET}$ would outperform $FLGP_{MRI}$ and $FLGP_{PET}$ due to the availability of multimodal data. However, the results indicate otherwise, with $FLGP_{MRI\_PET}$ achieving the lowest training and test accuracy among them. This performance gap highlights the limitations of FLGP in effectively handling multimodal data, as it struggles to fully exploit the complementary information from different slices and modalities. This is not surprising as FLGP was designed to only use one tree in an individual, which assumes every image is from the same slice index from a single modality.

The MTGP based methods shown in Table IV demonstrate the effectiveness of using multiple trees in each individual to learn from specific slices from each modality. $MTGP_{MRI}$ and $MTGP_{PET}$ significantly outperform their $FLGP_{MRI}$ and $FLGP_{PET}$ in training and test accuracy, confirming that using multiple trees enables better representation. The multi-

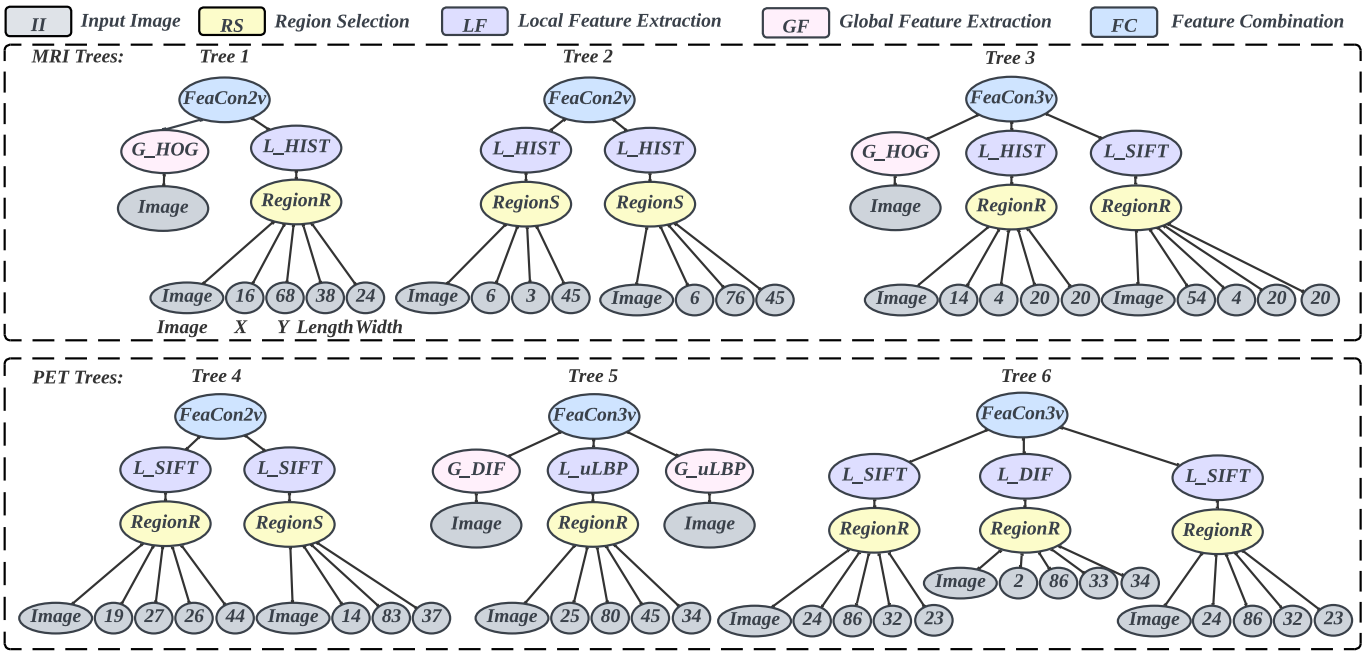

Fig. 4. An example program evolved by MMTGP on $Dataset_{MRI\_PET}$.

tree structure allows the model to explore diverse feature spaces while focusing on specific slices, enabling it to more effectively capture intricate patterns from corresponding brain regions.

Notably, MTGP_{PET} achieves a near-optimal training accuracy of 99.29%, highlighting the significant advantage of using multiple trees in GP to extract detailed functional features from PET data. When comparing MTGP based methods to MMTGP, MTGP trained on PET shows statistically better performance on training sets. However, MMTGP achieves statistically better performance on the test set, indicating that its multimodal design provides better generalization by effectively fusing complementary structural (MRI) and functional (PET) information. This result highlights the robustness and practical advantage of MMTGP in leveraging multimodal information for AD classification, outperforming single modality methods.

Demonstrated by Fig. 3, MTGP_{PET} and MMTGP achieve the highest positions in the plot, indicating their superior learning capabilities compared to the other methods. MTGP_{PET} excels due to its ability to capture detailed functional information from PET, while MMTGP effectively fuses complementary structural and functional information, enabling it to generalize well and maintain strong performance on both the training and test sets. Interestingly, in both plots, the curves involving PET data consistently achieve the highest positions. This suggests that when multimodal data is not available, or when cost and resource constraints are critical, PET may be a more effective and practical option for AD classification.

B. Program Analysis

The example individual program with six trees evolved by MMTGP is shown in Fig. 4, which achieves 99.12% accuracy on the training set and 100% accuracy on the test set with a total of 1117 features from six trees. To further evaluate the contribution of each modality, the features extracted by the MMTGP are tested using a linear SVM classifier. The modality 1 features are extracted from the MRI data, which achieves a test accuracy of 82.14%, while the features extracted from modality 2, the PET data, achieves a test accuracy of 96.43%. When the features from both modalities are combined, the test accuracy reached 100%, demonstrating the effectiveness of the proposed multi-tree structure in fusing complementary information from both modalities for optimal classification performance.

The example individual with six trees emphasizes both local and global feature diversity across the two modalities. Among the five local feature extraction methods, SIFT is the most frequently used, followed by HIST, with uLBP appearing once, while the remaining two methods are not utilized. For global feature extraction, HOG dominates, while SIFT is notably absent. This distribution indicates that different feature extraction methods are prioritized based on the type of information needed from the input data.

A significant observation is the difference in feature reliance between the two modalities. For the MRI modality, the model mainly depends on HIST and HOG, suggesting that MRI data primarily provides structural information, such as textures and gradients. This is exactly the kind of information that the MRI provides. On the other hand, PET relies heavily on SIFT, indicating that keypoint-based features, which capture localized functional changes, are more informative in PET scans. This distinction highlights the complementary nature of the two modalities: MRI focuses on capturing anatomical structures, while PET contributes functional details through localized keypoints. The ability to selectively extract modality-specific features underscores the importance of a multimodal approach, as it allows the model to combine structural and

functional information effectively for improved classification performance.

Moreover, different slices in different modalities present varying levels of complexity, and not all slices require both global and local feature extractions. From Fig. 4, Tree 2, Tree 4, and Tree 6 use only local feature extraction operators. Even when the same type of feature extraction operator is applied, variations can be observed. For instance, Tree 2 uses two HIST operators to extract features from two regions, while Tree 4 uses two SIFT operators for the same purpose, highlighting how different slices may require distinct types of local features to capture relevant information. Additionally, the two subtrees on the sides of Tree 6 are exactly the same. After removing the duplicated subtree, the training accuracy decreases to 96.44% but the test accuracy remained 100% accuracy. Besides, the features extracted from modality 2, the PET data, achieve a test accuracy of 100%. This indicates that the presence of redundancy requires further consideration for future improvements.

VI. Conclusions

This paper presents a new MMTGP approach for multimodal image classification, focusing on Alzheimer's Disease diagnosis using MRI and PET data. Using a multi-tree structure, MMTGP effectively captures complementary information from structural and functional modalities, significantly improving classification accuracy compared to single-modality models and traditional GP approaches. The results demonstrate that using multiple trees allows for diverse and robust feature extraction, enabling the model to explore intra-modality patterns and combinations.

The experimental results highlight that integrating modality-specific features plays a crucial role in enhancing performance, with MRI features primarily relying on texture and structural information (HIST and HOG), while PET data benefits from keypoint-based descriptors (SIFT). The ability to selectively extract and combine these features demonstrates the advantage of multimodal learning. Notably, MMTGP outperforms single-modality FLGP and MTGP on multimodal datasets, reinforcing the effectiveness of explicit multimodal feature fusion.

Although the multi-tree program structure improves feature representation, the gap between training and test accuracy, and the same feature extraction on the same region in some cases suggests the need for additional measures to mitigate the redundancy. Future research could investigate pruning strategies to reduce redundancy and improve the model's generalization.

References

[1] R. Brookmeyer, E. Johnson, K. Ziegler-Graham, and H. M. Arrighi, "Forecasting the global burden of alzheimer's disease," *Alzheimer's & dementia*, vol. 3, no. 3, pp. 186–191, 2007.

[2] M. A. DeTure and D. W. Dickson, "The neuropathological diagnosis of alzheimer's disease," *Molecular neurodegeneration*, vol. 14, no. 1, p. 32, 2019.

[3] J. R. Petrella, R. E. Coleman, and P. M. Doraiswamy, "Neuroimaging and early diagnosis of alzheimer disease: a look to the future," *Radiology*, vol. 226, no. 2, pp. 315–336, 2003.

[4] H. Zhou, L. He, B. Y. Chen, L. Shen, and Y. Zhang, "Multi-modal diagnosis of alzheimer's disease using interpretable graph convolutional networks," *IEEE Transactions on Medical Imaging*, 2024.

[5] C. R. Jack Jr, M. A. Bernstein, N. C. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P. J. Britson, J. L. Whitwell, C. Ward *et al.*, "The alzheimer's disease neuroimaging initiative (adni): Mri methods," *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 27, no. 4, pp. 685–691, 2008.

[6] 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.

[7] C. Choudhury, T. Goel, and M. Tanveer, "A coupled-gan architecture to fuse mri and pet image features for multi-stage classification of alzheimer's disease," *Information Fusion*, vol. 109, p. 102415, 2024.

[8] F. Krones, U. Marikkar, G. Parsons, A. Szmul, and A. Mahdi, "Review of multimodal machine learning approaches in healthcare," *Information Fusion*, vol. 114, p. 102690, 2025.

[9] Y. Bi, B. Xue, P. Mesejo, S. Cagnoni, and M. Zhang, "A survey on evolutionary computation for computer vision and image analysis: Past, present, and future trends," *IEEE Transactions on Evolutionary Computation*, vol. 27, no. 1, pp. 5–25, 2022.

[10] Q. Fan, Y. Bi, B. Xue, and M. Zhang, "Multi-tree genetic programming for learning color and multi-scale features in image classification," *IEEE Transactions on Evolutionary Computation*, 2024.

[11] A. Parziale, R. Senatore, A. Della Cioppa, and A. Marcelli, "Cartesian genetic programming for diagnosis of parkinson disease through handwriting analysis: Performance vs. interpretability issues," *Artificial intelligence in medicine*, vol. 111, p. 101984, 2021.

[12] Q. U. Ain, B. Xue, H. Al-Sahaf, and M. Zhang, "Multi-tree genetic programming with a new fitness function for melanoma detection," " in *IEEE Congress on Evolutionary Computation*, pp. 880–887, 2019.

[13] Y. Bi, B. Xue, and M. Zhang, "An automatic feature extraction approach to image classification using genetic programming," " in *Applications of Evolutionary Computation: 21st International Conference, EvoApplications 2018, Parma, Italy, April 4-6, 2018, Proceedings 21*, pp. 421–438, 2018.

[14] Q. U. Ain, H. Al-Sahaf, B. Xue, and M. Zhang, "Automatically diagnosing skin cancers from multimodality images using two-stage genetic programming," *IEEE Transactions on Cybernetics*, vol. 53, no. 5, pp. 2727–2740, 2022.

[15] S. Liu, S. Liu, W. Cai, H. Che, S. Pujol, R. Kikinis, D. Feng, M. J. Fulham *et al.*, "Multimodal neuroimaging feature learning for multiclass diagnosis of alzheimer's disease," *IEEE transactions on biomedical engineering*, vol. 62, no. 4, pp. 1132–1140, 2014.

[16] Y. Bi, B. Xue, and M. Zhang, "An effective feature learning approach using genetic programming with image descriptors for image classification [research frontier]," *IEEE Computational Intelligence Magazine*, vol. 15, no. 2, pp. 65–77, 2020.

[17] S. Tang, Q. Chen, B. Xue, M. Huang, and M. Zhang, "Genetic programming with multi-task feature selection for alzheimer's disease diagnosis," " in *IEEE Congress on Evolutionary Computation*, pp. 1–8, 2024.

[18] W. D. Penny, K. J. Friston, J. T. Ashburner, S. J. Kiebel, and T. E. Nichols, *Statistical parametric mapping: the analysis of functional brain images*. Elsevier, 2011.

[19] V. Fonov, A. C. Evans, K. Botteron, C. R. Almli, R. C. McKinstry, D. L. Collins, B. D. C. Group *et al.*, "Unbiased average age-appropriate atlases for pediatric studies," *Neuroimage*, vol. 54, no. 1, pp. 313–327, 2011.

[20] 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.

[21] 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*, pp. 1469–1472, 2010.

[22] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, pp. 91–110, 2004.

[23] Z. Yan, Y. Bi, B. Xue, and M. Zhang, "Automatically extracting features using genetic programming for low-quality fish image classification," " in *2021 IEEE Congress on Evolutionary Computation (CEC)*, pp. 2015–2022, 2021.