*Journal of Computer and Knowledge Engineering, Vol. , No.. 2023.*
*DOI:*

# GrAMIA-Net: Gray Matter Adaptive Multi-View Imbalance-Aware Attention Network for Detection of Parkinson's Disease

***Abstract*--** Diagnosing Parkinson's disease (PD) at its early stages from structural MRI remains a difficult task, primarily because of scanner-dependent variations, skewed class distributions, and the subtle anatomical alterations associated with early neurodegeneration. To improve MRI-based Parkinson's disease classification under these conditions, GrAMIA-Net is proposed as a multi-view deep learning framework built upon gray matter representations derived from T1-weighted MRI. The framework learns complementary information from sagittal, coronal, and axial perspectives and employs a hierarchical attention mechanism to emphasize discriminative spatial cues within each view. In addition, an inter-view attention module is utilized to aggregate information across different anatomical planes effectively. To mitigate the adverse effects of class imbalance, an Adaptive Imbalance Optimization (AIBO) scheme is integrated into the training process, thereby improving learning for underrepresented samples. Validation on the multi-scanner and class-imbalanced PPMI cohort shows that the proposed framework attains an accuracy of 0.862, an F1-score of 0.919, and an AUC of 0.720, demonstrating reliable predictive capability across categories. These results suggest that GrAMIA-Net provides a practical and computationally efficient framework for PD detection using single-modality GM representations, with potential applicability in real-world neuroimaging settings.

***Index Terms*--Adaptive Imbalance Optimization, Gray Matter Maps, Multi-View Learning, PD Detection, T1-weighted Structural Imaging.**

## 1. INTRODUCTION

Parkinson's disease (PD) is a chronic neurodegenerative condition and among the most prevalent movement disorders in the elderly population worldwide [1]. It is characterized by motor symptoms including tremor, muscular rigidity, and bradykinesia, as well as non-motor manifestations including cognitive and behavioral impairments [2–5]. Despite its clinical importance, PD diagnosis largely relies on subjective assessments, such as neurological examinations and rating scales (e.g., UPDRS) [6,7]. The absence of reliable objective biomarkers remains a major challenge for early diagnosis and highlights the need for automated and data-driven diagnostic approaches.

Neuroimaging has become an important tool for investigating structural brain alterations associated with PD. Among available imaging modalities, T1-weighted structural MRI is widely used because of its accessibility, relatively low cost, and ability to provide detailed anatomical information [8–10]. Although conventional MRI does not offer disease-specific visual markers, it enables quantitative assessment of structural changes in brain regions associated with PD progression, including the basal ganglia and cortical areas [11,12]. In particular, gray matter (GM) has attracted considerable attention, as GM atrophy and cortical thinning have been linked to disease severity and progression [13,14].

Despite these advantages, MRI-based PD analysis remains challenging. Variability introduced by different scanners and acquisition protocols can affect image characteristics, making standardized preprocessing procedures such as bias-field correction, skull stripping, and spatial normalization essential for ensuring data consistency [15–17]. Furthermore, GM representations often exhibit low contrast and complex anatomical structures, making subtle disease-related patterns difficult to identify. Another important challenge is class imbalance, which is common in PD datasets and may bias predictive models toward the majority class while reducing sensitivity to clinically relevant cases [18].

Data-driven learning methods, particularly CNN-based models, have shown considerable potential for MRI-based classification of PD by capturing relevant patterns in neuroimaging data [19–22]. Multi-branch architectures, ensemble strategies, and hybrid frameworks that combine deep learning with conventional machine learning techniques have further improved classification performance [23,24]. However, conventional CNNs are inherently limited in modeling long-range spatial dependencies and complex contextual relationships within neuroimaging data [25]. Although attention mechanisms have been introduced to alleviate these limitations [27–31], many existing methods rely on volumetric representations or simple feature fusion strategies and do not explicitly model hierarchical intra-view relationships or inter-view dependencies across sagittal, coronal, and axial planes.

Another challenge is the effective integration of information from multiple anatomical views. Common fusion approaches, such as feature concatenation and averaging, may not fully exploit the complementary information available across different views [32,33]. Additionally, methods that rely on multimodal imaging data or artificially balanced datasets often encounter limitations when applied to heterogeneous clinical data collected under real-world conditions.

Building on these observations, this study presents GrAMIA-Net, a multi-view framework that utilizes gray matter (GM) maps derived from T1-weighted MRI for Parkinson's

disease classification. The proposed framework learns complementary anatomical information from sagittal, coronal, and axial views using a hierarchical attention mechanism that captures multi-scale structural patterns in each view. To facilitate effective integration of information across anatomical planes, a dedicated cross-attention module is employed to model inter-view relationships and enhance feature representation. Additionally, an Adaptive Imbalance Optimization (AIBO) strategy is incorporated into the training process to enhance learning robustness in the presence of highly imbalanced data distributions.

By combining multi-view gray matter representation learning, hierarchical attention-based feature extraction, cross-view feature interaction, and imbalance-aware optimization, GrAMIA-Net provides an effective framework for automated Parkinson's disease detection in heterogeneous neuroimaging environments.

Unlike existing MRI-based PD classification methods that either rely on simple view fusion or process anatomical views independently, the proposed framework jointly addresses hierarchical intra-view representation learning, adaptive cross-view integration, and severe class imbalance within a unified gray matter–based architecture.

The key contributions of the present study are outlined below:

• A compact framework for GM-based Parkinson's disease detection: GrAMIA-Net is developed as a deep learning architecture that leverages multi-view gray matter representations derived from T1-weighted MRI for automated PD classification.

• Hierarchical intra-view attention modeling (HIA-GM): A dedicated attention module is introduced to capture multi-scale gray matter characteristics within sagittal, coronal, and axial planes, enabling structured feature learning at different anatomical scales.

• Cross-view anatomical feature integration (MVCA): The proposed MVCA module models interactions among anatomical views through cross-attention, allowing complementary information from different planes to be effectively aggregated into a unified representation.

• Adaptive imbalance optimization (AIBO): An imbalance-aware training strategy combining SMOTE-based oversampling, targeted augmentation, and class-weighted focal loss is designed to improve robustness and sensitivity when learning from highly imbalanced clinical datasets.

The rest of the manuscript is organized as follows. Section 2 outlines the materials and methodology. Section 3 presents the experimental findings, performance evaluation, interpretability analysis, and comparative studies. Section 4 summarizes the main conclusions and discusses future research directions.

## 2. Materials and Methods

The following section describes the proposed GrAMIA-Net for automated Parkinson's disease classification using GM maps derived from T1-weighted MRI. The framework integrates multi-view representation learning, attention-based feature modeling, and imbalance-aware optimization within a unified architecture.

### *A.Overview of the Proposed GrAMIA-Net Framework*

The Overall GrAMIA-Net architecture is provided in Fig. 1. The network receives sagittal, coronal, and axial GM maps generated through the preprocessing and segmentation pipeline. These views are fed into the network to learn view-specific features.

Intra-view features are modeled using the Hierarchical Intra-View Attention (HIA-GM) module, which combines convolutional feature extraction with spatial attention to capture multi-scale patterns. The Multi-View Cross Attention (MVCA) module models dependencies across anatomical views and enables effective feature fusion.

The resulting multi-view representation is then forwarded to a fully connected classification head for PD prediction. To address class imbalance, an imbalance-aware optimization strategy is applied during training, incorporating data augmentation, SMOTE-based oversampling, and class-weighted focal loss .The training process was conducted in two stages. In the first stage, the HIA-GM and MVCA modules were trained to learn discriminative multi-view feature representations from GM maps. In the second stage, the learned representations were utilized by the MLP (multi-layer perceptron) classifier for final PD prediction. This strategy facilitated stable feature learning while reducing the risk of overfitting on the limited training data.

Additional details regarding the individual components of GrAMIA-Net are provided in the following subsections.

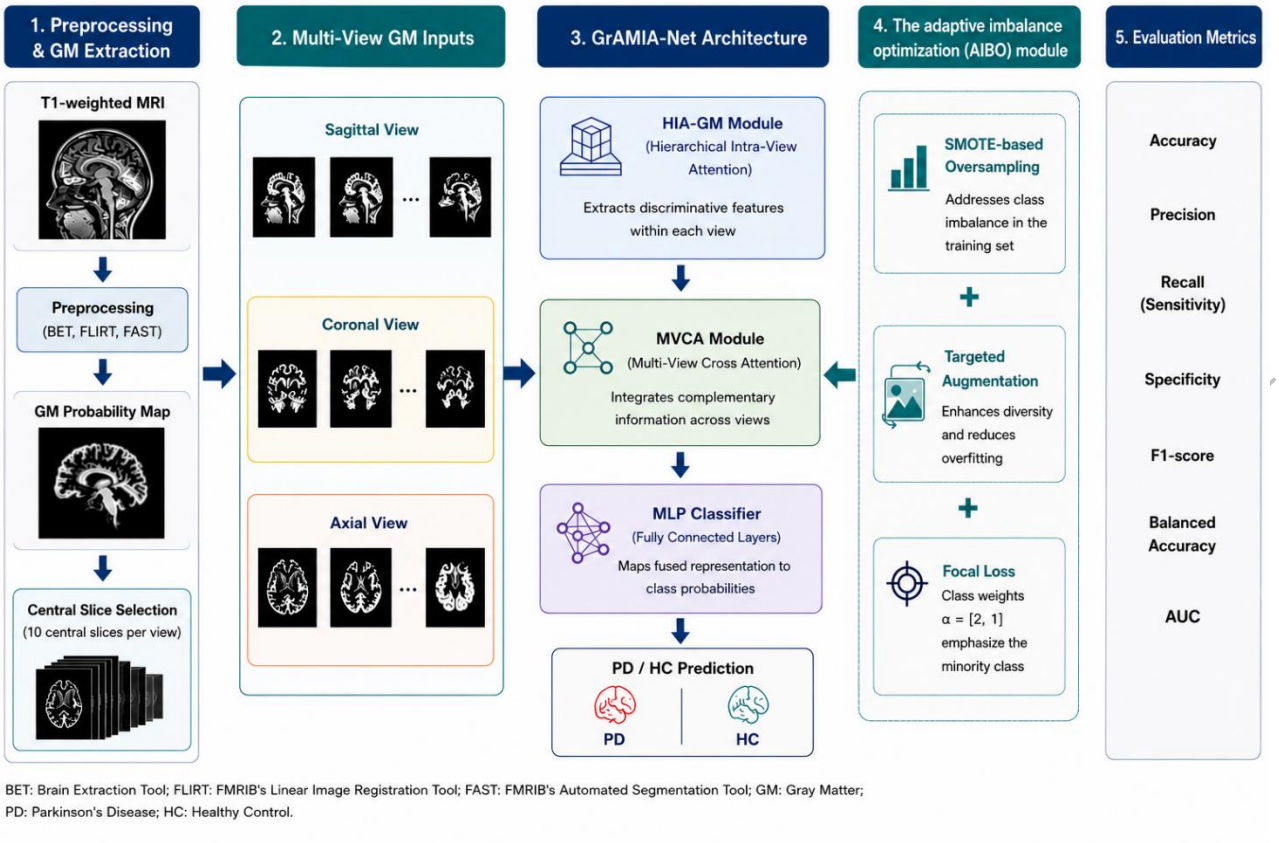



Fig. 1. The architecture of the proposed GrAMIA-Net framework.

*B. Data Acquisition*

The MRI data analyzed in this work were obtained from the Parkinson's Progression Markers Initiative (PPMI). This international multicenter study provides standardized clinical, imaging, and biomarker data for Parkinson's disease research [34]. Diagnostic labels were obtained directly from the PPMI database (www.ppmi-info.org) and were based on the official PPMI analytic cohort assignments. Only subjects belonging to the Parkinson's disease (PD) and Healthy Control (HC) cohorts were included in this study because the objective of the present work was to investigate binary PD classification. Participants assigned to other categories, such as SWEDD, prodromal, or genetic cohorts, were excluded because these groups represent distinct clinical populations with different diagnostic objectives and disease characteristics. Including such subjects would transform the task from binary PD classification into a substantially different multi-cohort classification problem. No additional relabeling or manual annotation was performed. Subject inclusion was determined exclusively by cohort assignment and MRI availability. No subjects were selected or excluded according to image appearance, disease severity, model predictions, or study outcomes, thereby reducing the risk of subjective selection bias during dataset construction.

The data were acquired using 3D T1-weighted MRI scans from 3.0 Tesla scanners, including models from Siemens, Philips, GE Medical Systems, and Philips Healthcare, with an image resolution of 256 × 256 × 192 voxels. Acquisition parameters, such as pulse sequences, flip angles, and repetition times (TR), differed across scanner platforms. A summary of the dataset characteristics is provided in Table 1.

TABLE 1. The demographic information of datasets

| Manufacturer of Scanners | Age | PD | Control | Sex |
|---|---|---|---|---|
| SIEMENS, Philips Medical Systems, GE MEDICAL (Flip Angle: [9,12], Pulse Sequence: GR, GR/IR, TR: [6.5,2300]) | Min: 34 Max: 86 | 288 | 37 | Female:111 Male:214 |

*C.MRI Preprocessing and Gray Matter Map Extraction*

To improve consistency across scans acquired from different imaging systems, all MRI volumes were processed using a standardized workflow implemented in the FMRIB Software Library (FSL) [35]. The preprocessing pipeline, summarized in Fig. 2, consisted of DICOM-to-NIfTI conversion, image intensity standardization, orientation alignment, bias-field correction, brain extraction, and spatial registration.

Brain tissues outside the cranial region were removed using the BET tool. The resulting images subsequently underwent

intensity inhomogeneity correction and affine registration to a common anatomical template using FLIRT. Tissue segmentation was performed with FAST to obtain probabilistic representations of GM, WM, and CSF. Finally, the GM representations were refined by applying thresholding and spatial smoothing operations to suppress noise and improve anatomical consistency across subjects.

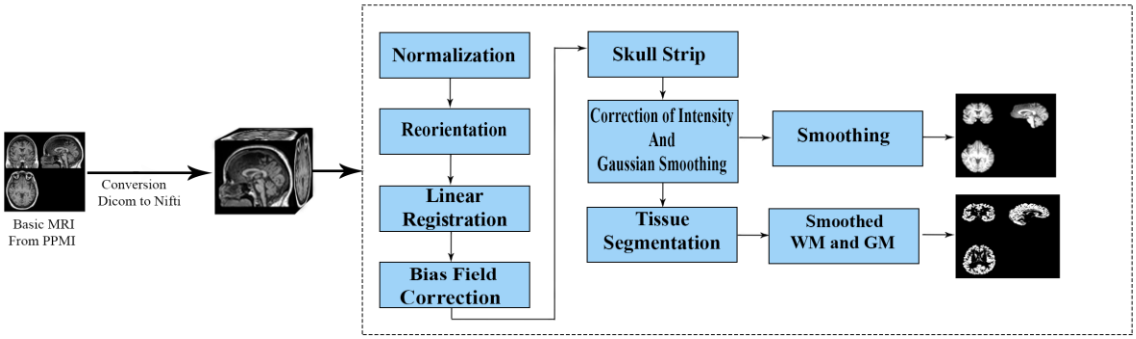



Fig. 2. The preprocessing pipeline.

Following preprocessing, 10 central slices were selected from each anatomical view (sagittal, coronal, and axial), as these regions typically contain more informative structural patterns relevant to neurodegeneration, while peripheral slices contribute less discriminative information [22]. This selection also reduces computational complexity while preserving diagnostically relevant features. The number of selected slices is consistent with related multi-plane MRI classification approaches [36], where central slice windows are commonly adopted as a practical and effective input strategy. Visual inspection of the preprocessed GM maps further confirmed that the selected region exhibited the highest gray matter density and the most pronounced structural variation across subjects.

Although alternative slice counts may also retain useful anatomical information, a fixed set of 10 central slices was adopted to balance anatomical coverage, computational efficiency, and model complexity across all experiments. Despite this reduced input representation, the proposed framework achieved competitive classification performance relative to previously reported MRI-based approaches (Table 5), suggesting that the selected slice configuration preserves sufficient discriminative information for PD classification while substantially reducing computational requirements.

*D.Hierarchical Intra-View Attention Module (HIA-GM)*

The Hierarchical Intra-View Attention Module (HIA-GM) is a core component of the proposed GrAMIA-Net, designed to model GM structures within each anatomical plane. As illustrated in Fig. 3, the module captures hierarchical spatial dependencies across sagittal, coronal, and axial views while preserving full spatial resolution.

For each view, input feature maps are processed independently through three sequential attention-guided convolutional stages. A sequence of 3 × 3 convolution, batch normalization, nonlinear activation, and spatial attention is employed at each stage.

This progressive design enhances the effective receptive field, allowing for the extraction of both local structural details and broader anatomical context without relying on pooling operations.

Spatial attention is computed using channel-wise average and max pooling, followed by convolution and sigmoid activation to generate attention maps. These maps are applied multiplicatively to the feature representations, while residual connections preserve low-level spatial information and support stable training.

The stacked attention-enhanced stages produce multiscale intra-view representations with increasing contextual awareness. These representations are aggregated to form refined, view-specific embeddings that encode structured anatomical information. By explicitly modeling intra-view spatial hierarchies, the HIA-GM module enables precise characterization of subtle GM alterations associated with early-stage PD and provides informative features for subsequent cross-view fusion.

*E.Multi-View Cross Attention (MVCA) Module*

Conventional multi-view fusion strategies, such as concatenation, averaging, or majority voting, combine features without explicitly modeling inter-view dependencies, often assuming equal contribution from each anatomical plane [36]. This can limit their ability to capture complementary and context-dependent relationships among sagittal, coronal, and axial representations.

To address this limitation, we introduce a Multi-View Cross Attention (MVCA) module that captures interactions between view-specific GM features. The module operates on latent representations $F \in \mathrm{R}^{n\times 3\times d}$, where $n$is the batch size, 3 denotes the anatomical views, and $d$is the feature dimension, obtained from the HIA-GM module. The input features are transformed into query (Q), key (K), and value (V) representations through trainable linear mappings. Cross-attention is computed via scaled dot-product attention [37]:

$$Attention(Q,K,V) = Softmax(QK^{\top})V \quad (1)$$

In the standard transformer formulation, a scaling factor of $(1/\sqrt{d})$ is commonly applied to stabilize the magnitude of dot-product attention for high-dimensional feature representations. In the proposed MVCA module, attention is computed over three compact view-level feature embeddings generated by HIA-GM. During model development, both the scaled and unscaled attention formulations were implemented and evaluated under the same experimental setting. The unscaled formulation consistently produced slightly better classification performance and was therefore adopted in the final architecture. This comparison was performed to guide the architectural design of the proposed model rather than as a dedicated ablation study. Accordingly, the present work does not claim that removing the scaling factor is generally advantageous, and this design choice should be interpreted only in the context of the proposed framework.

From others, enabling context-aware feature integration. The attention-enhanced features are then processed by a feedforward network (FFN) with nonlinear activation and dropout. A residual connection is applied by adding a linearly projected version of the input, followed by layer normalization:

$$Z=LayerNorm(Proj(F)+FFN(Attention(Q,K,V))) \quad (2)$$

where $Q,K,V \in R^{n\times 3\times d}$and $Z \in R^{n\times 3\times d}$. The output is reshaped to $R^{n\times 3d}$and passed to the classification head.
Unlike standard transformer architectures, the proposed MVCA is tailored for cross-view spatial fusion rather than sequential modeling. By enabling interactions across anatomically distinct yet correlated GM representations, it facilitates effective integration of complementary structural information. Fig. 4 provides an overview of the MVCA module.

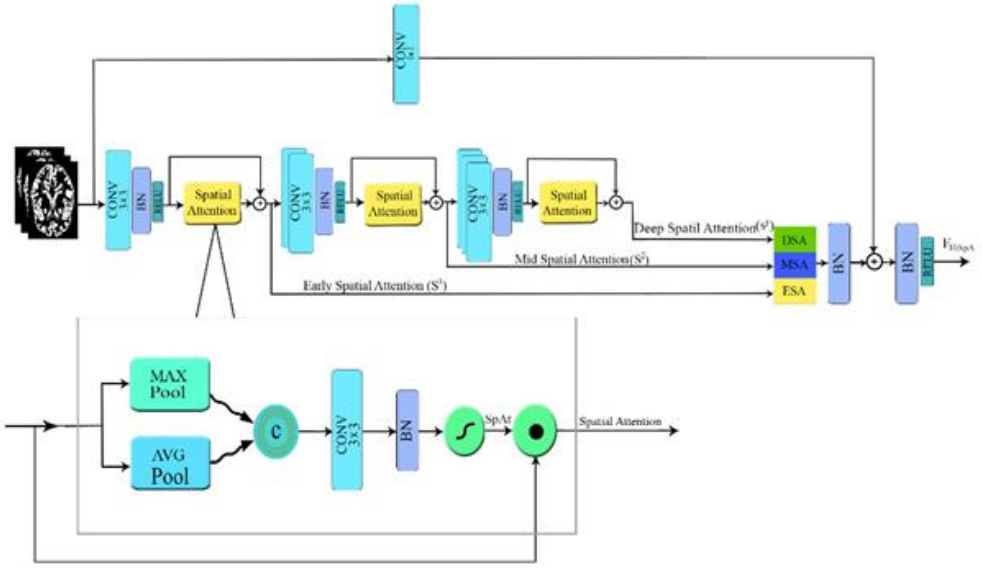



Fig. 3. The structure of the HIA-GM module.

*F. Final classification via MLP*

After multi-view fusion, the resulting feature representation is fed into a fully connected for classification. The MLP maps high-dimensional embeddings to diagnostic labels while maintaining computational efficiency. The classification head employs a progressively decreasing layer configuration, where the feature dimension is reduced from 256 units in the first layer to 32 units in the final hidden layer. ReLU activation, L2 regularization, and dropout (0.2) are applied after each layer to improve generalization. The final layer uses a softmax activation to produce probabilistic outputs for binary PD classification.

The classification head was optimized for 200 epochs using a batch size of 8. This classification head complements the preceding feature extraction and attention modules, supporting efficient end-to-end training while maintaining discriminative capability.

*G.The adaptive imbalance optimization (AIBO) module*

Class imbalance remains a major issue in medical imaging, often favoring dominant classes and limiting the detection of minority patterns in Parkinson's disease datasets [6]. To address this, the proposed framework incorporates an Adaptive Imbalance Optimization (AIBO) module that combines three complementary strategies:

SMOTE-based oversampling: SMOTE is employed during training to synthesize additional minority-class samples through interpolation, improving class balance and decision boundary generalization.

Class-weighted focal loss: To emphasize hard-to-classify samples, a class-weighted focal loss [38] is employed:

$$L_{\text{Focal}} = \frac{1}{N}\sum_{i=1}^{N} \alpha_i(1-\hat{y}_i)^{\gamma}[-y_i\, log(\hat{y}_i) - (1-y_i)\, log(1-\hat{y}_i)] \quad (3)$$

that $N$ is the batch size, $y_i \in \{0,1\}$is the ground truth, and $\hat{y}_i$is the predicted probability. The parameters $\alpha_i$and $\gamma$control class weighting and focus on hard samples, respectively.

Targeted data augmentation: Minority-class samples are

augmented using small rotations (±15°) and horizontal flips to simulate realistic variations without distorting anatomical structures.

By integrating synthetic sampling, adaptive loss weighting, and targeted augmentation, the AIBO module improves robustness and enhances sensitivity to minority-class patterns under imbalanced conditions.

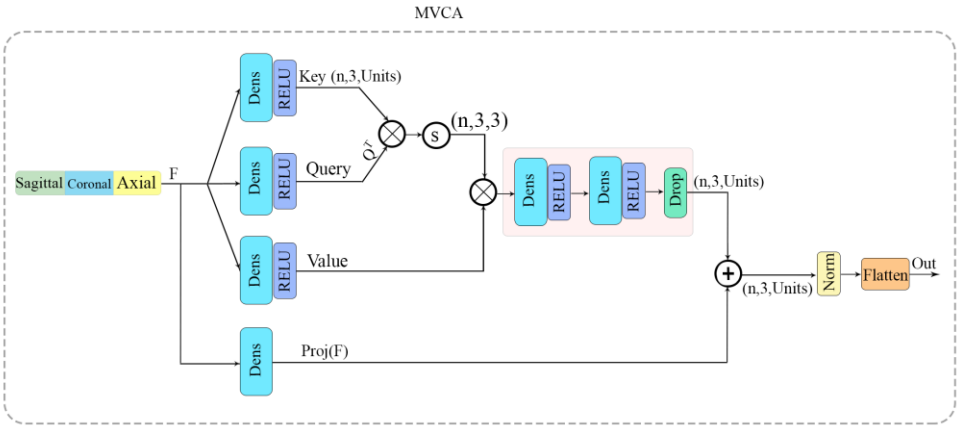



Fig. 4. The proposed MVCA module.

*H. Integration of HIA-GM and MVCA: Workflow and Complexity Analysis*

To improve the clarity of the proposed architecture, Fig. 5 illustrates the computational workflow of the GrAMIA-Net attention framework. The flowchart summarizes the interaction between the HIA-GM and MVCA modules, highlighting how hierarchical intra-view feature extraction and cross-view feature fusion are integrated to generate the final multi-view representation.

To provide additional insight into the computational characteristics of the proposed framework, the computational complexity of GrAMIA-Net was derived from the operations performed within the HIA-GM and MVCA modules illustrated in the flowchart. Let H×W denote the spatial resolution of the feature maps, V=3 the number of anatomical views, S=10 the number of input channels per view, and $C_1$ =43, $C_2$ =85, and $C_3$ =128 the numbers of output channels in the three hierarchical stages of HIA-GM.

Since each view is processed independently through three sequential convolutional-attention stages, the dominant computational cost of HIA-GM can be expressed as: $T\_HIA\text{-}GM = O(VH^2(SC_1 + C_1C_2 + C_2C_3)$. Using the actual architectural parameters of the proposed model, the computational cost of HIA-GM is approximately $2.71\times10^{10}$ multiply–accumulate operations (MACs) per subject. In contrast, the MVCA module operates on only three view-level feature representations after global pooling, resulting in a complexity of $O(Vd^2)$, where d denotes the feature dimension. With V=3 and d=256, the computational cost of MVCA is approximately $7.86\times10^5$ MACs. Therefore, the computational cost of GrAMIA-Net is dominated by the hierarchical feature extraction process of HIA-GM, while the additional overhead introduced by MVCA is negligible in comparison. This analytical observation was further supported by empirical runtime measurements on the held-out test cohort. Feature extraction required 2.34s, 2.29s, and 2.94s for the sagittal, coronal, and axial views, respectively.

The MVCA module required 0.03s, and the final MLP classifier required 0.05s for the complete test cohort. The total inference time was approximately 7.64s, corresponding to an average inference time of approximately 117ms per subject. These measurements confirm that the computational cost is primarily associated with feature extraction, whereas cross-view attention contributes only a small additional overhead. Furthermore, by processing a fixed set of ten informative central slices per anatomical view rather than the full MRI volume, the proposed framework substantially reduces memory consumption and computational requirements while maintaining competitive classification performance.

## 3. Experimental results and discussion

This section investigates the behavior of GrAMIA-Net and analyzes its performance in PD classification based on multi-view GM maps.

*A. Experimental setup and evaluation protocol*

The experimental evaluation was performed on a workstation featuring an AMD Radeon RX 5700 XT GPU and 64 GB of RAM. MRI preprocessing was performed using Nipype in conjunction with the FMRIB Software Library (FSL) on a Linux platform, while deep learning experiments were conducted in Python using TensorFlow/Keras. Following preprocessing and GM segmentation, 10 central slices were selected from each anatomical view (sagittal, coronal, and axial) as model inputs, as these regions typically contain more informative structural patterns.

Data partitioning was performed at the subject level, allocating 60% of the subjects to training, 20% to validation, and the remaining 20% to testing, ensuring no subject overlap across subsets and preventing data leakage. To improve learning under imbalanced conditions, SMOTE-based

oversampling and targeted augmentation were applied exclusively to the training set through the AIBO strategy, whereas the validation and test subsets were left unchanged during evaluation.

Training followed a two-stage procedure consistent with the architecture outlined in Section 2.1. During the initial training phase, the HIA-GM and MVCA backbone modules were trained end-to-end for 50 epochs using batches of eight samples, an initial learning rate of 0.01, and L2 regularization of 0.001. In the second stage, the MLP classification head was trained for 200 epochs using the same batch size and regularization settings. The focal loss parameters were empirically optimized with class weights $\alpha = [2, 1]$ to emphasize the minority class. Training parameters were selected based on validation-set performance and convergence behavior.

To assess training stability, all experiments were repeated ten times with independent weight initializations while preserving the same subject-wise train, validation, and test partitions. All metrics are presented as the mean values across repeated runs.

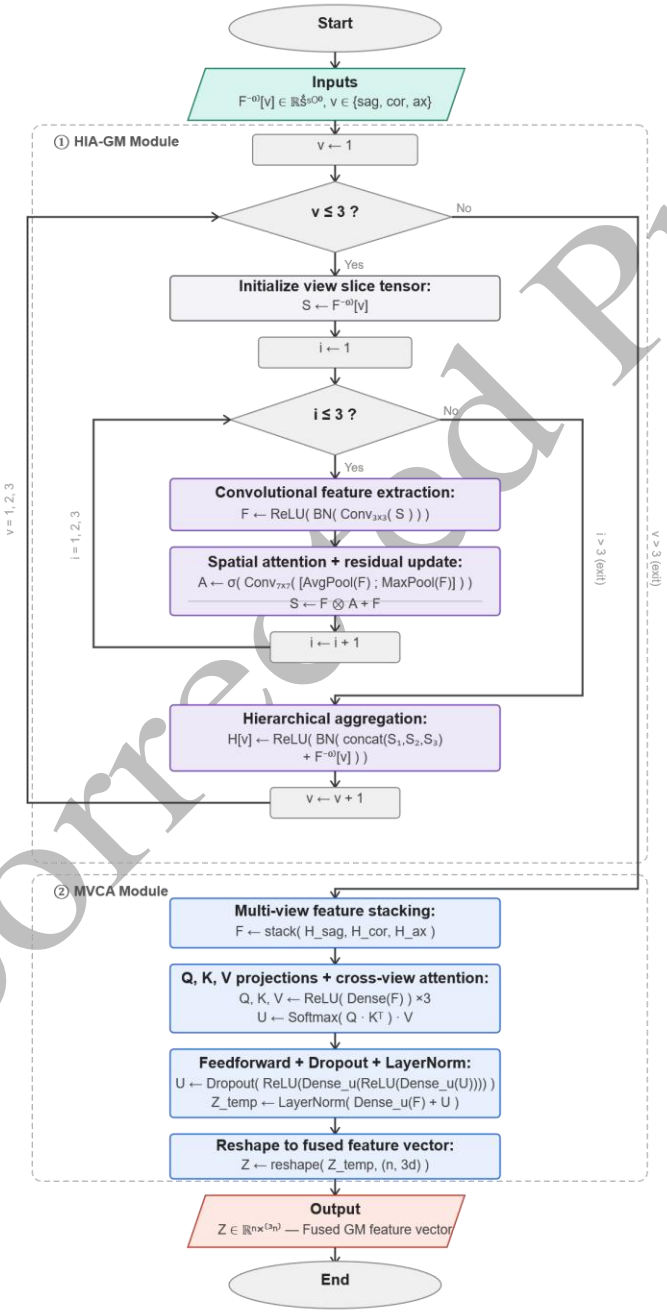



Fig. 5. Flowchart of the HIA-GM and MVCA module.

*B. Performance evaluation metrics*

Performance was assessed using commonly adopted classification measures, including specificity, accuracy (ACC), precision, sensitivity, F1-score, balanced accuracy (BACC), and area under the ROC curve (AUC), computed from the confusion matrix (TP, TN, FP, FN) [29].

Given the class imbalance, particular emphasis was placed

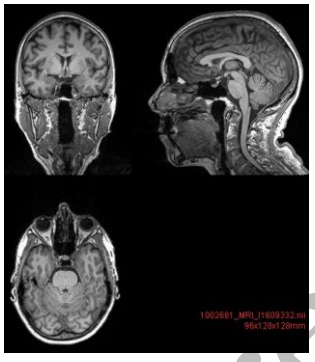
on F1-score, balanced accuracy, precision, and specificity, as accuracy alone may provide misleading estimates under skewed class distributions. These metrics offer complementary perspectives on model performance under class imbalance.

*C. Qualitative analysis of preprocessing*

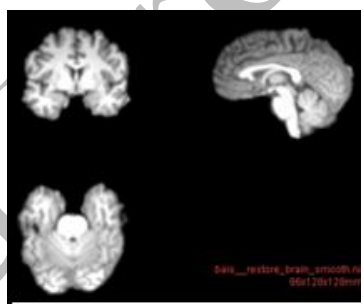
To assess the impact of the preprocessing pipeline, representative examples of processed T1-weighted images and their corresponding GM maps are shown in Fig. 6. The preprocessed images exhibit improved spatial consistency, effective removal of non-brain tissues, and reduced background noise compared to raw MRI scans.

In particular, GM maps provide clearer structural delineation and suppress non-informative regions, resulting in more focused and anatomically meaningful representations. Although the visual differences are subtle, they contribute to improved feature stability and support more effective downstream learning. These images and corresponding GM maps are shown in Fig.6. The pre-processed images exhibit improved spatial consistency, effective removal of non-brain tissues, and reduced background noise compared to raw MRI scans.

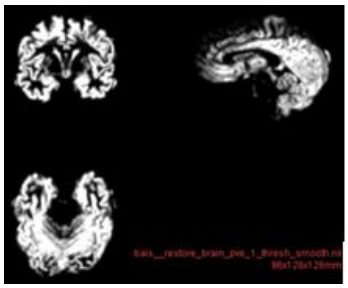
In particular, GM maps provide clearer structural delineation and suppress non-informative regions, resulting in more focused and anatomically meaningful representations. Although the visual differences are subtle, they contribute to improved feature stability and support more effective downstream learning. These observations are consistent with prior studies highlighting the importance of intensity normalization and spatial registration in neuroimaging-based analysis [16,17].



(a): Sample of MRI scans in nifti format from 3 views (coronal, sagittal, and axial)



(b): Sample of T1-MRI scans after preprocessing steps



(c): Sample of GM maps after preprocessing steps

Fig. 6. The preprocessing pipeline outcomes.

*D. Quantitative performance evaluation*

The quantitative performance of GrAMIA-Net is presented in Fig. 7, where metric values are reported as the average of ten independent experimental runs. An F1-score of 0.92 was obtained, suggesting effective performance under imbalanced conditions. Precision remained high (0.93) with minimal variability, supporting reliable detection of PD cases. Recall and specificity showed moderate fluctuations across runs, possibly influenced by class imbalance and scanner variability. Despite this variation, balanced accuracy (0.73) and AUC (0.72) indicate stable overall classification performance.

Although specificity remained lower than sensitivity, this behavior is likely a consequence of the severe class imbalance of the dataset. Additional evidence supporting this interpretation is provided by the ablation results presented in the Imbalance Handling subsection in Section 3.F, where the removal of individual imbalance-aware components leads to substantial reductions in specificity and balanced accuracy.

From a practical perspective, these results indicate that GrAMIA-Net maintains reliable discrimination between PD and HC subjects despite the challenges posed by severe class imbalance and scanner heterogeneity. The observed balanced accuracy suggests that the model does not merely favor the majority class but retains meaningful predictive capability across both diagnostic groups. Furthermore, the combination of high precision and F1-score indicates that the proposed framework can identify PD cases with a relatively low rate of false positive predictions while preserving strong overall classification performance. Collectively, these findings support the effectiveness of the proposed multi-view attention

framework for Parkinson's disease classification in realistic clinical settings.

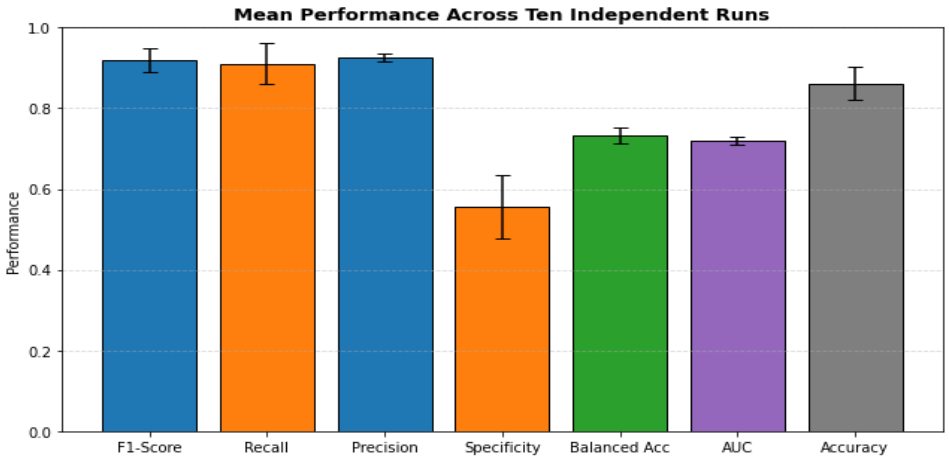



Fig. 7. Mean classification performance of GrAMIA-Net across ten independent runs.

To evaluate the impact of prediction granularity, the model was assessed at both slice-level and subject-level resolutions. In the slice-level setting, individual slices were processed independently while maintaining subject-wise separation to prevent data leakage. As reported in Table 2, slice-level evaluation yields lower performance (F1-score = 0.73, AUC = 0.62). Although precision remains high, reduced recall indicates increased sensitivity to noise and local variations.

In contrast, subject-level aggregation improves performance (F1-score=0.92, AUC=0.72), suggesting that integrating information across slices enhances global anatomical representation and reduces variability. While alternative slice-level settings may increase recall, they often reduce specificity and yield an unstable AUC, indicating less reliable predictions. Overall, subject-level inference provides more robust outcomes for PD classification.

Table 2. Comparison of slice-level and subject-level performance

| Evaluation Level | Slice-Level | Subject-Level (Proposed) |
|---|---|---|
| Accuracy | 0.615 | 0.862 |
| Precision | 0.919 | 0.927 |
| Recall | 0.607 | 0.911 |
| F1-score | 0.731 | 0.919 |
| Specificity | 0.667 | 0.556 |
| Balanced Acc | 0.637 | 0.733 |
| AUC | 0.621 | 0.720 |

The contribution of multi-view learning was further evaluated using individual anatomical views and their combined representation (Fig. 9). Single-view models show inferior performance, with AUC values below 0.62. Among them, the axial view performs best (F1-score=0.79), and then coronal (0.76) and sagittal (0.68). In contrast, the multi-view configuration achieves higher performance (F1-score =0.92, AUC = 0.72), highlighting the benefit of integrating complementary anatomical information across views

Finally, the impact of input modality was examined by comparing raw T1-weighted MRI with GM maps. GM-based inputs improve performance (F1-score: 0.73 → 0.92; AUC: 0.65 → 0.72), indicating that GM representations provide more discriminative structural information by reducing irrelevant variability.

### *E. Attention Map Visualization*

To interpret cross-view interactions, average attention weights were analyzed for both training and test sets (Fig. 10). The attention matrices exhibit consistent patterns across both sets, indicating stable and generalizable feature integration. Across all query views, higher attention is assigned to axial representations, followed by coronal, while sagittal contributions are comparatively lower. This suggests that the model prioritizes complementary structural information from axial and coronal planes during feature fusion. From a neuroimaging perspective, these findings are consistent with the distributed nature of PD-related structural alterations reported in the literature [39]. However, the reported attention weights reflect the relative importance of sagittal, coronal, and axial representations during feature fusion rather than the spatial localization of disease-related anatomical changes.

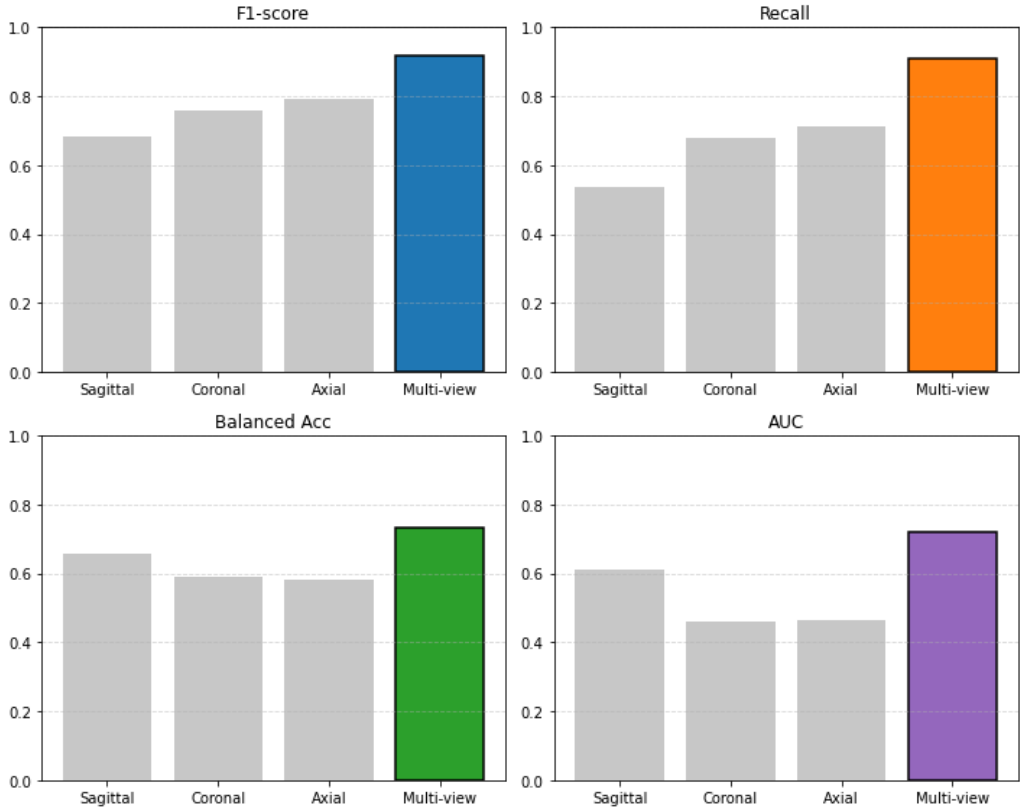



Fig. 9. Performance comparison of single-view and multi-view configurations.

Overall, the attention patterns indicate effective modeling of cross-view dependencies and adaptive weighting of view-specific features, supporting both interpretability and robustness. Consequently, the presented analysis provides architectural interpretability by revealing how information is weighted across anatomical views during feature fusion, but it should not be interpreted as direct anatomical attribution. Localization-based explainability methods, such as Grad-CAM, could provide complementary spatial information and may further enhance the interpretability of the proposed framework.

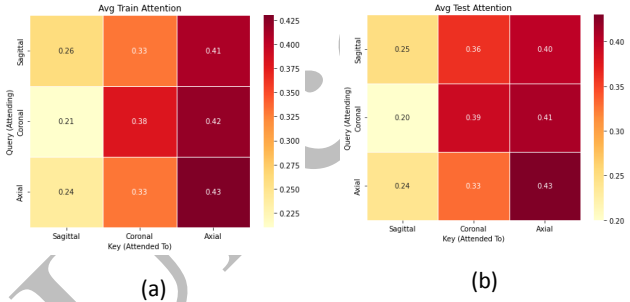



Fig. 10. Average cross-view attention matrices for training and test data.

*F. Ablation and Comparative Analysis*

The effects of the architectural modules, fusion scheme, and imbalance optimization strategy were investigated through ablation studies.

**Core components:** As shown in Table 3, removing either HIA-GM or MVCA leads to a noticeable performance drop, while removing both results in the largest degradation, confirming their complementary roles in intra-view feature extraction and cross-view integration.

Table 3. Performance impact of core Architectural Components in the GrAMIA-Net.

| Architecture Variant | Accuracy | Balanced ACC | AUC | F1-Score | ΔAUC |
|---|---|---|---|---|---|
| GrAMIA-Net (Full Model) | 0.862 | 0.733 | 0.720 | 0.919 | --- |
| Without spatial attention (HIA-GM) | 0.616 | 0.684 | 0.655 | 0.725 | -0.065 |
| Without MVCA | 0.646 | 0.655 | 0.663 | 0.757 | -0.058 |
| Without MVCA & HIA-GM (Basic CNN) | 0.569 | 0.610 | 0.557 | 0.689 | -0.163 |

**Fusion strategy:** Fig. 11 shows that MVCA achieves better performance than concatenation and majority voting, highlighting the benefit of modeling interactions across anatomical views. In contrast, concatenation fails to capture cross-view relationships, while majority voting produces less stable predictions.

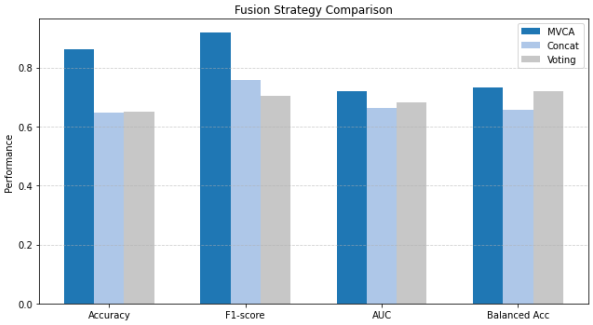



Fig. 11. Comparison of fusion strategies.

**Imbalance handling:** The impact of AIBO is presented in Table 4. The full configuration achieves the most balanced performance. Removing SMOTE substantially decreases specificity and balanced accuracy, indicating an increased tendency toward majority-class predictions. A comparable reduction is observed when data augmentation is removed, suggesting that augmentation also contributes to improving minority-class recognition and overall model robustness. In contrast, removing focal loss primarily reduces recall and F1-score, highlighting the complementary roles of the individual components of the proposed AIBO strategy in addressing severe class imbalance. Although specificity remained lower than sensitivity, this behavior is consistent with the severe class imbalance of the cohort, where PD subjects substantially outnumber healthy controls. Nevertheless, the full AIBO configuration achieved a balanced accuracy of 0.733, indicating improved balance between sensitivity and specificity under highly imbalanced conditions.

Table 4. Effect of AIBO components on model performance

| Configuration | Accuracy | Balanced Acc | F1-Score | Recall | Precision | Specificity |
|---|---|---|---|---|---|---|
| AIBO (SMOTE + Focal + Aug) | 0.86 | 0.73 | 0.92 | 0.92 | 0.93 | 0.56 |
| Without SMOTE | 0.83 | 0.58 | 0.90 | 0.91 | 0.88 | 0.22 |
| Without Focal Loss | 0.68 | 0.67 | 0.78 | 0.67 | 0.92 | 0.56 |
| Without Augmentation | 0.74 | 0.52 | 0.84 | 0.82 | 0.86 | 0.22 |

*G. Comparison with existing methods*

Table 5 summarizes the performance of GrAMIA-Net alongside recent MRI-based methods for Parkinson's disease classification. As these studies differ in input modality, dataset size, and class distribution, the comparison should be interpreted as indicative rather than strictly controlled.

GrAMIA-Net achieves competitive performance using only GM maps derived from T1-weighted MRI, a widely available clinical imaging modality. This demonstrates that competitive performance can be achieved using a single MRI modality, without requiring multimodal imaging inputs. Compared with conventional GM-based machine learning approaches, such as VBM+SVM [13] and Multi-feature ML [20], the proposed framework achieves substantially higher classification performance. This improvement may be attributed to the combination of hierarchical attention-based representation learning and structured multi-view feature integration, which enable the model to capture complementary anatomical information across views. Deep learning approaches based on larger cohorts, such as the 3D CNN model reported in [22], achieve higher AUC values. Similarly, multimodal architectures, including SCTransformer-VGG11 [36], benefit from the integration of complementary imaging modalities. However, these approaches either rely on substantially larger datasets or require additional imaging information beyond GM representations. In contrast, GrAMIA-Net was evaluated on a highly imbalanced cohort representative of real-world clinical conditions. Furthermore, the proposed framework outperforms earlier GM-map-based approaches, including the Ensemble ML + 3D CNN model [40], while maintaining a relatively simple imaging input pipeline. The contribution of the AIBO module is reflected in the balanced accuracy and F1-score values, indicating improved robustness across classes despite the severe class imbalance of the dataset.

Overall, these findings suggest that GrAMIA-Net provides an effective compromise between performance, robustness, and model complexity in realistic multi-scanner settings.

*H. Generalization and limitations*

GrAMIA-Net demonstrates stable generalization on a multi-scanner, highly imbalanced dataset that reflects realistic clinical conditions. A subject-wise partitioning strategy was used to maintain strict independence among the training, validation, and test sets. Despite relying solely on GM maps derived from T1-weighted MRI, the model achieves competitive performance, indicating that discriminative structural features can be effectively captured using a single, widely available modality. The multi-view design and attention-based fusion further enhance robustness to scanner variability.

However, several limitations should be noted. The dataset exhibits substantial class imbalance, which may still affect model calibration. The cohort's severe imbalance also remains a potential source of bias and should be further investigated using larger and more balanced datasets. In addition, the study was validated using only one dataset, potentially restricting generalizability to other cohorts. Although GrAMIA-Net was assessed on a heterogeneous multi-scanner cohort, external validation on independent datasets was not performed. Accordingly, the results provide evidence of within-dataset generalization rather than definitive proof of cross-cohort robustness. In addition, the number of selected slices was fixed across all experiments. Although this choice was motivated by anatomical relevance, prior literature, and computational efficiency considerations, a systematic sensitivity analysis across alternative slice counts was not performed. Therefore, the optimal number of slices remains an open question and should be investigated in future studies.

Another limitation of the proposed framework is its higher

computational complexity compared with conventional single-view models. GrAMIA-Net processes three anatomical views and incorporates hierarchical attention-based feature extraction together with cross-view feature fusion, which increases the computational cost of the framework. This additional complexity represents the trade-off for learning richer multi-view representations and improving robustness to scanner variability. Nevertheless, the framework operates on compact gray-matter slice representations rather than full volumetric MRI data, helping to keep the computational requirements manageable. As discussed in the Computational Complexity Analysis (Section 2.H), the computational cost is primarily dominated by the HIA-GM feature extraction stage, whereas the additional overhead introduced by the MVCA module is relatively small. Empirical runtime measurements further showed that the complete inference process required approximately 117ms per subject on the held-out test cohort, demonstrating that the proposed framework remains computationally practical despite its increased architectural complexity.

Furthermore, GM-based representations depend on preprocessing pipelines that may introduce variability. While attention mechanisms improve interpretability, their direct clinical relevance remains to be validated. Future studies will investigate external validation using independent and multi-center datasets, the incorporation of longitudinal imaging data, systematic evaluation of alternative slice-selection strategies, and the development of more clinically grounded interpretability frameworks.

Table 5. Comparison of the MAVI-Net model with existing methods.

| Method | Data type | Subjects (PD/HC) | Accuracy | F1-score | AUC | Balanced Acc |
|---|---|---|---|---|---|---|
| Ensemble ML + 3D CNN [40] | GM Maps | 187/187 | 0.620 | — | 0.630 | — |
| Multi-feature ML (GM+WM) [20] | GM+WM | 177/67 | 0.842 | 0.848 | — | — |
| 3D CNN [22] | T1-MRI | 1024/1017 | 0.793 | — | 0.87 | — |
| ICA + 3D CNN (SCTransformer-VGG11)[36] | T1+fMRI | 127 / 36 | 0.840 | — | 0.850 | — |
| Feature-based + Decorrelated CNN [6] | rs-fMRI | 164/234 | 0.805 | 0.772 | 0.623 | 0.667 |
| VBM + SVM [13] | GM + demo | 88/120 | 0.670 | — | 0.630 | — |
| GrAMIA-Net (Proposed) | GM Maps | 288/37 | 0.862 | 0.919 | 0.720 | 0.733 |

## 4. Conclusion and Future Work

This paper presented GrAMIA-Net, a multi-view DL framework for PD classification based on GM maps extracted from T1-weighted MRI. The framework combines hierarchical intra-view attention with cross-view feature fusion to capture structural information from sagittal, coronal, and axial representations. In addition, the proposed AIBO strategy improves learning under imbalanced data distributions. The experimental findings indicate that GrAMIA-Net achieves competitive performance using only GM-based inputs derived from a single imaging modality. The integration of attention-guided feature learning and multi-view representation modeling contributes to robust feature extraction and effective disease classification. These results suggest that GM representations can provide informative biomarkers for automated PD detection in realistic neuroimaging settings.

Future research will investigate the generalizability of the framework using independent multi-center cohorts. Additional directions include incorporating complementary imaging modalities and clinical information, as well as further improving the interpretability and robustness of the proposed architecture.

## REFERENCES

[1] M. Absalyamova, Traktirov, D., Burdinskaya, V., Artemova, V., Muruzheva, Z., & Karpenko, M . "Molecular basis of the development of Parkinson's disease." *Neuroscience*. pp. 292-300,2025.

[2] N. Vijiaratnam, T. Simuni, O. Bandmann, H. R. Morris, and T. Foltynie, "Progress towards therapies for disease

modification in Parkinson's disease." *The Lancet Neurology* vol. 20 no. 7, pp. 559-572,2021.

[3] Zhang, X., Zhang, L., Xu, K. et al. "Optimal channel selection based on ABC-SVM in Parkinson′s disease detection using short-time resting state EEG." *J Supercomput* 81, 977, 2025.

[4] S. J. Duncan, Marques, K., Fawkes, J., Smith, L. J., & Wilkinson, D. T., "Galvanic vestibular stimulation modulates EEG markers of voluntary movement in Parkinson's disease." *Neuroscience* vol. 555, pp. 178-183,2024.

[5] X. Liu, Li, W., Liu, Z. et al, "A dual-branch model for diagnosis of Parkinson's disease based on the independent and joint features of the left and right gait," *Appl Intell*, 2021.

[6] Patil, P., & Ford, W. R. "Parkinson's Disease Recognition Using Decorrelated Convolutional Neural Networks: Addressing Imbalance and Scanner Bias in rs-fMRI Data." *Biosensors*, 2024.

[7] S. G. Ryman and K. L. Poston, "MRI biomarkers of motor and non-motor symptoms in Parkinson's disease." (in eng) *Parkinsonism & related disorders*, 2020.

[8] A. S. Talai, J. Sedlacik, K. Boelmans, and N. D. Forkert (2021) "Utility of Multi-Modal MRI for Differentiating of Parkinson's Disease and Progressive Supranuclear Palsy Using Machine Learning." (in eng) *Frontiers in neurology*, 2021.

[9] E. Porter, A. A. Roussakis, N. P. Lao-Kaim, and P. Piccini, "Multimodal dopamine transporter (DAT) imaging and magnetic resonance imaging (MRI) to characterise early Parkinson's disease",(in eng) *Parkinsonism & related disorders* 79, 26–33,2020.

[10] B. Hazhirkarzar et al., "Machine Learning for Prediction of Parkinson's Disease Progression Using DAT-SPECT Imaging and Multimodal MRI." *Journal of Nuclear Medicine* vol. 65, 2024.

[11] Y. Zhou, S. Tinaz, and H. D. Tagare, "Robust Bayesian Analysis of Early-Stage Parkinson's Disease Progression Using DaTscan Images." *IEEE Transactions on Medical Imaging* vol. 40 no. 2, pp. 549-561, 2021.

[12] E. Adeli et al., "Joint feature-sample selection and robust diagnosis of Parkinson's disease from MRI data." (in eng) *Neuroimage*, 2016.

[13] Anupa A. Vijayakumari, Hubert H. Fernandez & Benjamin L. Walter, "MRI‐based multivariate gray matter volumetric distance for predicting motor symptom progression in Parkinson's disease." *Scientific Reports*, 2023.

[14] Mohsen, M., Mohamed, N.A.E., Bedir, A.E.T.M. *et al.* Advanced MRI-based evaluation of gray and white matter changes in Parkinson's disease. *Egypt J Radiol Nucl Med* **55**, 165 (2024). https://doi.org/10.1186/s43055-024-01340-w

[15] Payares-Garcia, D., Mateu, J., & Schick, W. "NeuroNorm: An R package to standardize multiple structural MRI." *Neurocomputing*, 550, 126493, 2023.

[16] F. J. Martinez-Murcia, J. M. Górriz, J. Ramírez, and A. Ortiz, "Convolutional neural networks for neuroimaging in Parkinson's disease: Is preprocessing needed? " *International journal of neural systems*, Vol. 28, no. 10, 2018.

[17] S. Soltaninejad, I. Cheng, and A. Basu, "Towards the Identification of Parkinson's Disease Using only T1 MR Images." *in Smart Multimedia Springer International Publishing*, pp 145-156, 2018.

[18] R. Kumari, Singh, J., Gosain, A "Diagnosis of Parkinson Disease Using Ensemble Methods for Class Imbalance Problem." *Springer, Singapore. Proceedings of Data Analytics and Management. ICDAM 2023. Lecture Notes in Networks and Systems*, 2024.

[19] S. J. MacEachern and N. D. Forkert, "Machine learning for precision medicine." *(in eng) Genome*, pp 416–425, 2021.

[20] Sahu, S. K., & Chowdhury, A. S, "Multi-modal multi-class Parkinson disease classification using CNN and decision level fusion." *ArXiv*. 2023.

[21] D. Jin et al., "Attention-based 3D Convolutional Network for Alzheimer's Disease Diagnosis and Biomarkers Exploration. " *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, Venice, Italy, 2019.

[22] M. Camacho et al, "Explainable classification of Parkinson's disease using deep learning trained on a large multi-center database of T1-weighted MRI datasets." *NeuroImage: Clinical*, vol. 38. 2023.

[23] S. Soltaninejad, "Intelligent Parkinson's Disease Classification and Progress Monitoring using Non-invasive Techniques." *Doctor of Philosophy, Department of Computing Science, University of Alberta*, 2020.

[24] G. Solana-Lavalle, R. Rosas-Romero, "Classification of PPMI MRI scans with voxel-based morphometry and machine learning to assist in the diagnosis of Parkinson's disease." *Computer Methods and Programs in Biomedicine*, 2021.

[25] H. W. Loh et al., "Application of Deep Learning Models for Automated Identification of Parkinson's Disease: A Review (2011–2021)," *Sensors*, vol. 21, no. 21.

[26] Kinyanjui, T. Njoroge, Mugoye, K., and Kibuku, R. "Multi-Head Self-Attention Fusion Network for Enhanced Multi-Class Crop Disease Classification." *Journal of AI and Data Mining*, 13(2), pp.227-240, 2025.

[27] M. Golovanevsky, C. Eickhoff, and R. Singh, "Multimodal attention-based deep learning for Alzheimer's disease diagnosis." Journal of the American Medical Informatics Association, vol. 29, no. 12, 2022.

[28] Chauhan, V., Tiwari, A. & Kumar, A, "An attention mechanism-based hybrid TimeAttentionBiLSTM architecture for long-term traffic forecasting." *J Supercomput* 81, 1257. 2025.

[29] Y. Dai, Y. Gao, and F. Liu, "TransMed: Transformers Advance Multi-Modal Medical Image Classification." *Diagnostics*, 2021.

[30] Liu, Z., Lu, H., Pan, X., Xu, M., Lan, R., & Luo, X, "Diagnosis of Alzheimer's disease via an attention-based multi-scale convolutional neural network." *Knowledge-Based Systems*, 2022.

[31] Y. Niu, Chen, N., Zhu, H., Li, G., & Chen, Y, "Brain connectivity and time-frequency fusion-based auditory spatial attention detection." *Neuroscience*, vol. 560, pp. 397-405, 2024.

[32] Z. Liu, H. Lu, X. Pan, M. Xu, R. Lan, and X. Luo, "Diagnosis of Alzheimer's disease via an attention-based multi-scale convolutional neural network." *Knowledge-Based Systems*, vol. 238, 2022.

[33] S. Datta, M. K. Acharaya, B. Chakrabarty, and N. Ghosh, "MRI Voxel-Based Morphometry to Classify Brain Structural Alterations in Parkinson's Disease." *IEEE 21st India Council International Conference (INDICON)*, Kharagpur, India, 2024.

[34] K. Marek, D. Jennings, S. Lasch, A. Siderowf, C. Tanner, T. Simuni, et al., "The Parkinson Progression Marker Initiative (PPMI)," *Progress in Neurobiology*, vol. 95, no. 4, pp. 629–635, 2011.

[35] L. Dular, F. Pernuš, and Ž. Špiclin, "Extensive T1-weighted MRI preprocessing improves generalizability of deep brain age prediction models." *Computers in Biology and Medicine*, vol. 173, 2024.

[36] J. Liu, H. Du, Q. Bi, H. Liao, and Y. Pan (2022) "MEST: Multi-plane Embedding and Spatial-temporal Transformer for Parkinson's disease diagnosis." *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2022.

[37] K. Aghajani, "Deep Learning Approach for Robust Voice Activity Detection: Integrating CNN and Self-Attention with Multi-Resolution MFCC," *Journal of AI and Data Mining*, 2024.

[38] J. M. B. Md Sazzad Hossain, Andrew P. Paplinski, "Dual Focal Loss to address class imbalance in semantic segmentation." *Neurocomputing*. 2021.

[39] Xiaorui Cui, Lan Li, Lei Yu, Huijuan Xing, Hong Chang, Li Zhao, Jin Qian, Qingwei Song, Shiyu Zhou, Chunbo Dong (2020) "Gray Matter Atrophy in Parkinson's Disease and the Parkinsonian Variant of Multiple System Atrophy: A Combined ROI- and Voxel-Based Morphometric Study." *Clinics*, 2020.

[40] Islam, N.U., Khanam, R, "Machine learning-based Parkinson's disease detection and progression evaluation using gray and white matter segmentation from 3D MRI." *J. Eng. Appl. Sci*. 2024.