IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 25 April 2023, accepted 8 May 2023, date of publication 16 May 2023, date of current version 24 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3276722

RESEARCH ARTICLE

Same-Subject-Modalities-Interactions: A Novel Framework for MRI and PET Multi-Modality Fusion for Alzheimer's Disease Classification

BOUCHRA GUELIB[1], KARIM ZAROUR[1], HAITHEM HERMESSI[2], BOUNAB RAYENE[1], AND KHLIFA NAWRES[3]

[1]LIRE Laboratory, Faculty of New Technologies of Information and Communication, University of Abdelhamid Mehri Constantine 2, Constantine 25016, Algeria
[2]Laboratory of Informatics, Modeling and Information and Knowledge Processing (LIMTIC), Intelligent Systems in Imaging and Artificial Vision (SIIVA) Team, Higher Institute of Computer Science, University of Tunis El Manar, Ariana 2080, Tunisia
[3]Research Laboratory of Biophysics and Medical Technologies, Higher Institute of Medical Technologies of Tunis, University of Tunis El Manar, Tunis 1006, Tunisia

Corresponding author: Bouchra Guelib (bouchra.guelib@univ-constantine2.dz)

This work was supported in part by the PRFU Project C00L07UN250220200002.

This work involved human subjects or animals in its research. The author(s) confirm(s) that all human/animal subject research procedures and protocols are exempt from review board approval.

ABSTRACT Alzheimer's disease is a growing concern, and neuroimaging techniques such as Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET) scans are widely used to classify AD patients. While MRI captures structural information and measures brain atrophy, PET shows functional changes associated with neurological disorders, and both modalities have been proven to be AD biomarkers. However, combining MRI and PET in the same test without considering their inherent structural differences can result in a loss of important information. To address this issue, this paper proposes a novel machine learning framework for combining MRI and PET modalities and a new set of interactions known as Same-Subject-Modalities-Interactions (SSMI) to extract complementary information and new insights. The SSMI relation is derived from MRI and PET and subjected to PCA to construct the SSMI set, which is then concatenated with the other sets. The best set of features is selected and used for classification using Ridg-Classifier. Freesurfer is used to extract measures from 183 ADNI subjects (69 in the AD group and 114 in the CN group), and different classifiers are performed with train-test-split, cross-validation, and validation-set from ADNI-2/GO. The results showed high accuracy, precision, specificity, recall, F1-score, and AUC, with values of 98.94%, 98.27%, 97.10%, 100%, 99.13 and 98.55%, respectively, from ADNI and 98.75%, 98.48%, 93.75%, 100.0%, 99.23%, 96.80% from ADNI2/Go. These results are higher than those achieved by single-modality classification tasks and state-of-the-art approaches. Furthermore, the regions selected by Ridge Classifier are shown to be highly related to Alzheimer's disease biomarkers.

INDEX TERMS Multimodality, MRI, PET, fusion, machine learning, feature selection.

I. INTRODUCTION

Researchers are endeavouring to decelerate the symptoms of Alzheimer's disease (AD), provide a cure to patients, and prevent the onset of the disease. Additionally, they aim to decrease the cost of goods and services required for treating, caring for, and diagnosing patients. Neuroimaging data has been a valuable resource for researchers to achieve these goals. Due to the essential information, it provides about the brain, these techniques have shifted from being minor components to central components of AD diagnostics [1].

The associate editor coordinating the review of this manuscript and approving it for publication was Ines Domingues.

Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET) are commonly used in AD investigations for their structural and functional information. While MRI provides structural information and measures brain atrophy, PET shows metabolic quantity and neurotransmitter activity, which indicates functional changes associated with AD [2]. Therefore, both MRI and PET are considered as strongly related to AD biomarkers based on the characteristics they display in scan sessions.

However, despite the remarkable accuracy of using a single data type as input for machine learning (ML) models in the AD classification task, they fail to deal with complex data, extract new insights, or discover knowledge. Additionally, physicians typically search for atrophy associated with AD using specific locations known as ''regions of interest,'' which prevents the information from being utilized outside these regions. Furthermore, a complex phenomenon like disease, characterized by its rich nature, is rarely understood by a single modality alone [3]. In the prognosis process, physicians use various modalities to pinpoint the disease, such as questions in the initial interrogation to determine disease symptoms; in some cases, imaging scanners are required to confirm the diagnosis, which means that an accurate diagnosis requires various modalities to draw a holistic picture of the disease. The different modalities can be seen as puzzle pieces; by combining a global view and disease patterns, an overall picture can be established. Multimodality fusion combines different neuroimaging techniques to improve prognosis accuracy and gain new insights [4].

The fusion process is divided into three stages based on the fusion stage. Early fusion consists of concatenating the features of multiple modalities into a single vector or matrix. Therefore, complementary information is extracted, but heterogeneous information is destroyed since the features will be treated equally in the preprocessing step. The intermediate stage involves extracting and selecting features from each modality separately, then combining them. Thus, heterogeneity's nature is preserved, but the interactions between modalities that extract complementary information from modalities remain unexploited. The late stage entails constructing an intermediate model from each modality and fusing them to build the final model [5].

In recent decades, several studies [6], [7], [8], [9], and [10] used various modalities, including MRI, PET, Cerebrospinal Fluid (CSF), Computed Tomography Angiography (CTA), and Single-photon Emission Computerized Tomography (SPECT) to improve the model's accuracy, discover new knowledge, identify biomarkers, understand disease mechanisms, and establish a global and generic pattern for the disease. New searches demonstrate that combining MRI with other modalities like PET, demographic data, and cognitive scores can enhance the model's accuracy and reduce noise. On the other hand, the developed framework will suit real-world medical applications [11]. Furthermore, when local structure information is fully utilized and interactions across multiple modalities are taken into account, the classification's accuracy significantly improves [12]. Despite the benefits of multimodality fusion, various challenges and problems revolve mainly around data quality and the fusion process. The neuroimaging data are heterogeneous, complex, large in size, noisy, and high-dimensional, requiring sophisticated and developed techniques to handle these problems. AI and its techniques, namely ML and deep learning (DL), provide a paradigm shift to multimodality fusion powered by data availability and analytics techniques. ML techniques have demonstrated their robustness in dealing with data quality issues [13], [14]. In general, the state-of-the-art assesses the quality of proposed fusion approaches solely on the basis of model accuracy. One modality has already achieved high accuracy [15], [16]. However, the primary purpose of data fusion is to gain new knowledge, like discovering the gene potential related to AD. This goal is only possible by entirely using structure information across modalities, as proved in [17]. Interactions are also guaranteed, including inter- and intramodal relations as well as correlations between subjects of the same modality. This paper presents a novel fusion method based on three criteria: inter-relations, intra-relations, and same-subject-modalities-interactions. Additionally, to select and fuse features extracted from MRI and PET modalities to classify AD patients from the ADNI dataset. This paper is organized as follows: Section II focuses on background information. Section III contains materials and details the methodology. Section IV presents experimental results. Section V goes over the results and discusses the limitations. Section VII provides the final remarks and future works.

II. RELATED WORKS

Multimodality data fusion with ML techniques has been extensively studied in the research community. Several methods for combining different modalities at various stages have been proposed. Depending on the objectives of the analysis, ML techniques can be applied at any stage of the fusion process. The most straightforward approach is early fusion, which merges normalized features from different modalities into a single pool for classification. Early fusion methods such as SVM-based kernels, stepwise logistic regression analysis, and Gaussian processes have been used to model the interactions between the modalities and extract complementary information. However, early fusion has limitations in dealing with heterogeneous data, as it tends to prioritize one modality with a large number of features while discarding the information from other modalities.

Intermediate fusion is a different strategy that selects features independently for each modality, constructs a kernel-based matrix for each modality, and then combines them into a fused matrix. Similarity matrices can also be computed using Random Forest and used in the fusion process. However, these techniques can be sensitive to the weighting of different modalities, and assigning weights without careful consideration can lead to suboptimal performance.

Late fusion, on the other hand, treats each modality separately and disregards both inter and intra-modal interactions. Late fusion techniques cannot capture complementary information between modalities and are generally less effective than early or intermediate fusion methods.

This section provides a comprehensive review of various ML techniques for multimodal data fusion and highlights their advantages and disadvantages.

A. EARLY FUSION

For this task, the Logistic Regression Analysis (LRA), Support Vector Machine Classifier (SVM), and Gaussian Process (GP) are used. Several approaches have been proposed to fuse multimodality data for AD diagnosis at early stage. One approach by [18] concatenated features from MRI, fluorodeoxyglucose-PET (PET-FDG), and Cerebrospinal fluid (CSF) directly into a vector using multi-stepwise LRA, but did not exploit intra-relations or structure information. Another study by [19] used SVM to fuse VIO information from FDG-PET and MRI, without feature selection or complementary information extraction. Reference [20] selected features using ANOVA test and Principal component analysis (PCA), respectively, and evaluated performance using Gaussian classifier. However, both approaches combined features into a single pool, affecting modality heterogeneity and ignoring intra-relations.

B. INTERMEDIATE FUSION

Numerous previous studies applied Multiple kernel learning (MKL) to fuse selected features. MKL is also used to compute each modality's weight/coefficient and embed the fused weights into an SVM or another classifier. Random Forest (RF) is also used to compute similarity matrices to construct a unified matrix. GP, multi-view, and multi-task learning are applied in the fusion process. MKL is also used to compute each modality's similarity matrix or weight and embed the fused matrices/weights into an SVM or another classifier. RF is also used to compute MKL-SVM as a method for computing similarity matrix. Reference [21] suggested a framework for classifying AD using three imaging modalities: MRI, PET, and CSF, by computing a separate similarity matrix for each modality and then fusing them to construct a final matrix that can be integrated into the SVM classifier. However, combining the matrices without considering the heterogeneous nature of features may destroy them, and inter-relations are overlooked. To overcome this limitation, the study presented by [22] addressed AD vs. mild cognitive impairment tasks using the Gaussian process's multimodality fusion, which automatically learns each modality parameter, resulting in a balanced accuracy of 74%. Furthermore, [23] proposed a multi-task multimodal framework for AD classification by considering each modality as a separate task and using multi-task learning feature selection to select relevant features from each modality, which is then fused by a multimodal-based SVM classifier. The study in [24], addressed the need for an inter-modality relationship using multi-task learning by treating feature selection from each modality as a separate task and imposing an $l_{2,1}$-norm as a sparse constraint to preserve the inter-modality relations. Reference [25] proposed a fusion framework based on multi-task learning, where different learning tasks imposed a group sparsity regularizer to jointly select a subset of relevant features from various modalities. The study in [17], proposed a method for feature selection that fully uses structure information by computing each modality's similarity using RF, selecting features by constructing the objective function based on two regularization terms: group sparse and sample similarity constraint, and using MKL-SVM to fuse the selected features.

C. LATE FUSION

Late fusion or decision fusion involves training a classification model independently for each feature set/modality and fusing the different outcomes (classifier scores) into a final model. The standard method for combining different classifiers in late fusion is to compute a weighted total of the intermediate classifiers. Reference [26] proposed a sparse representation (SR) framework to classify AD patients using pre-selected features and constructing an intermediate classifier for each modality. A final model was constructed based on the mean residuals of the intermediate classifiers. Reference [27] used the same classifier, SRC, extended to weighted SRC, to fuse three modalities, selecting relevant features using the t-test method and assuming weights via greedy search. The issue of losing information characteristics during weight selection was addressed. Reference [28] fused kernel and Dartel using Simple MKL process and applied kernel ridge regression to predict subjects' age, diagnosis, and psychological scores. The kernel combination raised similar issues as [27]'s method.

D. DISCUSSION OF THE STATE-OF-THE-ART

Multimodal fusion using ML techniques has been extensively studied in previous research. However, while many studies have focused on exploiting inter-relations in intermediate fusion or leveraging complementary information in early fusion, the literature has largely overlooked the importance of examining the interactions between the same features of the same subject across different modalities. This limitation raises concerns about the validity of fusion results, as ignoring this relation may lead to the loss of crucial information and hinder accurate classification. Therefore, a robust multimodal fusion process should meet several criteria.

In addition to the established criteria for a robust multimodal fusion process, it is essential to assess the potential advantages and disadvantages of each technique with respect to computational complexity, interpretability, and generalizability. While they merge features from several modalities into a single feature vector, early fusion techniques are typically less computationally demanding. Unfortunately, this

might lead to information loss due to modality heterogeneity, and the resulting feature vector may be hard to interpret. On the other hand, late fusion techniques tend to have a larger computational complexity because they need training a distinct model for each modality and then integrating the outputs. The separate models can be examined to determine the contribution of each modality to the final choice, allowing for more interpretability.

Depending on the specific methodology, intermediate fusion methods can find a balance between computational complexity and interpretability. For instance, approaches that generate similarity matrices for each modality, may have high computational complexity, whereas alternatives that use multi-task learning for feature selection may be more interpretable.

Regarding generalizability, each approach has its own advantages and disadvantages. Initial fusion strategies may struggle with modality heterogeneity and may not generalize well to fresh datasets including multiple modalities. On the other hand, late fusion techniques may be more resilient to modality heterogeneity and easier to apply to new datasets. However, the performance of each method will depend on the dataset and application in question.

Based on prior work, we propose that a robust multimodal fusion process should meet the following criteria:

- Complementary information from modalities should be exploited by ensuring inter-relations between modalities.
- Structure information across modalities should be preserved by maintaining the heterogeneous nature of the data.
- Information from the same modality should be exploited through intra-relations.

In this paper, we present a novel framework for multimodal MRI and PET fusion that addresses important criteria that have been overlooked in previous studies. Our proposed approach considers the inter and intra-modality interactions between PET and MRI to extract complementary information for AD classification while preserving the local structure information inherent in each modality. Additionally, we introduce a new relation, the Same-Subject-Modalities-Interactions (SSMI), to generate a new set of data that enriches the training set and improves learning accuracy. To fully leverage the potential of the SSMI interaction, we propose a process based on Principal Component Analysis (PCA) to extract the most competent features that capture the maximum amount of variation in a given region based on the values captured by MRI and PET. Finally, we apply feature selection on the fused set to minimize the computational cost of the framework and extract relevant biomarkers.

Preprocessing of PET and MRI images is a crucial step in multimodality fusion for disease diagnosis and monitoring. PET images are registered to their corresponding MRI to enable segmentation of specific regions. This segmentation allows for the extraction of meaningful features that aid in disease classification. The alignment of PET with the anatomical information provided by MRI enables the identification of specific regions of interest. In addition to intra/inter-modality fusion, the SSMI interaction is important to consider. SSMI investigates how the same region of the same subject is captured by two different techniques and how these techniques can interact with each other. Investigating this interaction can provide complementary and pertinent information that may improve the accuracy of disease classification. To address this, our approach involves:

- Process and normalize each modality independently using Freesurfer to preserve the heterogeneous structure of each data source.
- Construct a new set based on the relations between different modalities for the same region and subject.
- Fuse each set in order to improve the learning performance.
- Perform RidgeClassifier (RC) for feature selection task to eliminate the variables with zero effect on the prediction and ensure inter-relations between modalities.
- Train multiple classifiers to achieve the best accuracy in the classification task.

III. MATERIALS AND METHODS

Data preparation, multimodality fusion, and classification are detailed in the following subsections.

A. DATA PREPARATION

Data preparation includes data description, MRI/PET preprocessing, and feature extraction.

1) DATASET DESCRIPTION

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ ADNI_Acknowledgement_List.pdf. The ADNI was launched in 2003 as a public-private partnership led by Principal Investigator Michael W. Weiner, MD. The primary goal of (ADNI) has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early AD. In this study, two imaging modalities, MRI and PET, are fused to conduct the classification task. A total of 366 participants were classified into the classes of MRI and PET. All the participants have both MRI and FDG-PET/ scans. Table 1 displays the considered subjects' demographic information.

2) MRI/PET ACQUISITION

We obtained 184 3D T1-weighted MRI scans in NiFTI format from the ADNI screening and baseline visits. These scans were acquired using a 3D sagittal volumetric

TABLE 1. Demographic information of the subjects considered.

Classes	AD	CN
Number of subjects	69	114
Female/Male	31/38	58/56
Age(year)AD	76.24 ±7.17	76.20±5.25

magnetization prepared rapid gradient echo (MP-RAGE) sequence and were already reviewed for quality. To correct for spatial distortion caused by gradient non-linearity and B1 field inhomogeneity, the scans underwent an automatic correction process followed by the N3 histogram peak-sharpening algorithm to reduce intensity and non-uniformity. We also downloaded 183 PET scans captured during the same period, which were acquired 30-60 minutes after injection and underwent preprocessing steps including co-registration, averaging, standardization, and resolution uniformity. Further details on the acquisition and preprocessing procedures are available on the ADNI website.

The data used in this study was obtained from the AD Neuroimaging Initiative (ADNI) Screening and ADNI Baseline. While the availability of data from this specific source is limited due to the requirement of having both MRI and PET scans for each patient, the advantage of using data from these visits is that they capture the earliest stages of cognitive decline and disease progression. By using data from these early visits, researchers can better understand the onset and progression of cognitive decline and AD pathology. This may allow for the development of more accurate diagnostic and prognostic models, which are critical in managing cognitive decline and improving patients outcomes.

3) MRI PREPROCESSING AND FEATURE EXTRACTION

Under Linux, we utilized Freesurfer [29] version 7.1.1, an open-source neuroimaging package, to preprocess and extract features from MRI and PET scans. Freesurfer consolidates several processing steps into a single command (recon-all) and is commonly used for skull-stripping, normalization, subcortical and white matter segmentation, and cortical parcellation of T1-weighted standard MNI 305 space MRI scans. Due to the time-consuming nature of the recon-all command, we employed the parallel command to process multiple scans simultaneously.

The recon-all command enables segmentation of the MRI brain into various regions and provides statistical information about each region. We utilized these statistics to create features in four steps:

a: STEP1: SUBCORTICAL SEGMENTATION

we used the *aseg.stats* files generated by Freesurfer to extract information on 45 subcortical brain regions. For each region, we extracted three measures: Normalized Deviation *NormDev*, NormMean value *NormMean*, and the volume of regions *volume*. We then concatenated the names of the regions with their measures to create a feature for each region. For example, the feature *Left-Lateral-Ventricle-Volume* represents the Left-Lateral-Ventricle region with the measure of volume, and its value is the volume of the Left-Lateral-Ventricle. In total, we extracted 135 features from this step (45 regions with three measures).

b: STEP2: CORTICAL PARCELLATION

We extracted 68 cortical regions' names (34 for the left cortex and 34 for the right) along with eight measures from *lh.aparc.stats*. These measures include the surface area *SurfArea*, the volume of the gray matter *GrayVol*, the average thickness of the cortex *ThickAvg*, the standard deviation of the cortical thickness across the vertices within a specific region *ThickStd*, the curvature of a surface *MeanCurv*, the Gaussian curvature *GausCurv*, the folding index of the brain's cortical surface *FoldInd*, and the curvature of the cortex surface *CurvInd*. Additionally, we extracted the white matter volume *WM* of the cortex's regions from *wmparc.stats*. The features were created by concatenating the regions' names with their corresponding measures. In total, we extracted 612 features (68 regions with nine measures).

c: STEP3: eTIV NORMALIZATION

The volumetric measures obtained from *aseg.stats* were normalized using the values of eTIV (Estimated Total Intracranial Volume) [20], [30]. eTIV refers to the total volume of the cranial cavity, including the brain, cerebrospinal fluid, and blood vessels. Neuroimaging techniques such as MRI are commonly used to measure eTIV, which is important in neuroimaging research because variations in head size can affect the accuracy and consistency of brain structure and function measurements. Normalizing volumetric measures with eTIV accounts for individual differences in head size, resulting in more accurate and reliable measurements of brain structure and function. In AD research, eTIV is often used as a covariate to control for differences in head size and brain atrophy among individuals. Some studies suggest that eTIV may be a useful predictor of AD risk and is associated with cognitive performance [31].

d: STEP4: MRI FEATURES

The subcortical and cortical regions were concatenated to construct the final MRI dataset for the upcoming experiments, resulting in a total of 747(135 subcortical features and 612 cortical feature) features.

4) PET PREPROCESSING AND FEATURE EXTRACTION

As we used Freesurfer for MRI processing, we also utilized it to process PET scans and extract the Standardized Uptake Value (SUV) as the measure for feature extraction. SUV [32] is a quantitative measure used in PET imaging to assess the metabolic activity in tissues. It represents the concentration of a radiopharmaceutical in a specific tissue relative to the injected dose and the body weight of the individual being imaged. This process involves four steps:

a: STEP1: PET REGISTRATION

Registration is a critical step in PET processing to extract as much information as possible from the low-resolution PET scans and to provide the same segmentation as MRI. This step is crucial to modeling new relationships between MRI and PET regions, as explained in Section III-B3. The *bbregister* command in Freesurfer utilizes *SPM* as a registration tool, providing reliable PET registration to T1-weighted MRI scans. The registration step results in multiple files, and the file with the *lta* extension is used for volume mapping.

b: STEP2: PET VOLUME MAPPING

The volume mapping step plays a crucial role in PET processing by resampling the PET volume into the same field of view as the MRI using various matrices (such as FreeSurfer, FSL, SPM, and MNI). To achieve this, the *mri_vol2vol* command is applied using the SPM registration file, resulting in an output image with an *mgh* extension. This mapped PET image will be used in the subsequent segmentation step.

c: STEP3: PET SEGMENTATION

The PET segmentation step involves dividing the PET region into 45 subcortical and 68 cortical regions, similar to the MRI segmentation process. To achieve this, the *mri_segstats* command was utilized with the *aparc+aseg.mgz* file, which provided the same parcellation of MRI-PET as previously established in [33] and implemented in [34]. The resulting output from PET segmentation is a *dat* file, which includes the mean intensity of each FreeSurfer region (ROI).

d: STEP4: SUV EXTRACTION AND PET SET CONSTRUCTION

The features of PET are its 113 regions (45 subcortical and 68 cortical) and their SUV value, which were obtained after the PET segmentation step. The PET set is a *csv* file with 113 regions and their corresponding SUV values.

Figure 1 shows the complete pipeline of MRI and PET Freesurfer processing.

B. MULTIMODALITY FUSION FRAMEWORK

The proposed framework consists of four steps: MRI and PET preprocessing and feature extraction, multimodality fusion, feature selection, and classification. The first step is explained in Section III-A, and the proposed architecture of multimodality fusion is depicted in Figure 2.

Previous works on multimodality fusion did not consider the inherent knowledge in each modality and ignored the intra/inter interactions among and between modalities, resulting in poor complementary information extraction. However, the proposed framework aims to preserve the nature structure of information in each modality and extract complementary information between modalities by ensuring intra/inter interactions. To achieve this, the framework proposes four criteria: structure information, intra-relations, SSMI, and inter-relations, which are detailed in the following subsections.

1) STRUCTURE INFORMATION ACROSS DIFFERENT MODALITIES

Local structure information refers to the distinctive characteristics and topology of the data in each modality, as noted by [17], [35], and [36]. In the context of fusing MRI and PET scans, this refers to the anatomical information that can be obtained from these modalities, including the shape, size, and spatial location of different structures and regions within the body. It is essential to preserve the initial heterogeneity of the data as it contains valuable information that can improve the accuracy and robustness of the analysis.

However, combining data from different sources in the same pool may destroy this initial heterogeneity, which pertains to the inherent differences between the data from distinct sources, such as variations in data modalities, acquisition protocols, and patient populations. To preserve the structure information of each modality, we propose using different preprocessing steps for each modality separately. This approach ensures that the unique nature of each modality is preserved, given that applying the same preprocessing step to all data may destroy their uniqueness.

In our proposed work, we used Freesurfer to process each modality separately, and the features that require normalization are normalized independently. For instance, in MRI, the region volume, gray matter volume, and white matter volume are normalized using eTIV https://surfer.nmr.mgh.harvard.edu/fswiki/eTIV), which is defined as the volume within the cranium, including the brain, meninges, and CSF' [37]. On the other hand, in PET, the SUV is normalized by the standardized uptake value ratios, which is the whole cerebellum SUV (the sum of subcortical and cortical SUV). The SUVs are computed by dividing the SUV by the SUVr. By normalizing the MRI and PET features independently, their unique structure information is preserved.

2) INTRA-RELATIONS

Intra-relations refer to the relationships, correlations, and connections between the regions of the same modality within each subject. The importance of these interactions can be seen when performing a single-modality analysis. The relations between the regions of the same modality reveal standard information that can be used to generalize knowledge about the disease and the modality used in the analysis. To extract information within a single modality in a signal classification task, the intra-relations between the features within that modality are critical. These interactions between the features of the same modality reveal valuable insights that contribute to the accuracy and robustness of the analysis.

To ensure these interactions are preserved, all subcortical and cortical regions are extracted from each modality and concatenated into the same training set. During feature selection, the selection method considers all the features to select the relevant ones, further emphasizing the importance of intra-relations.

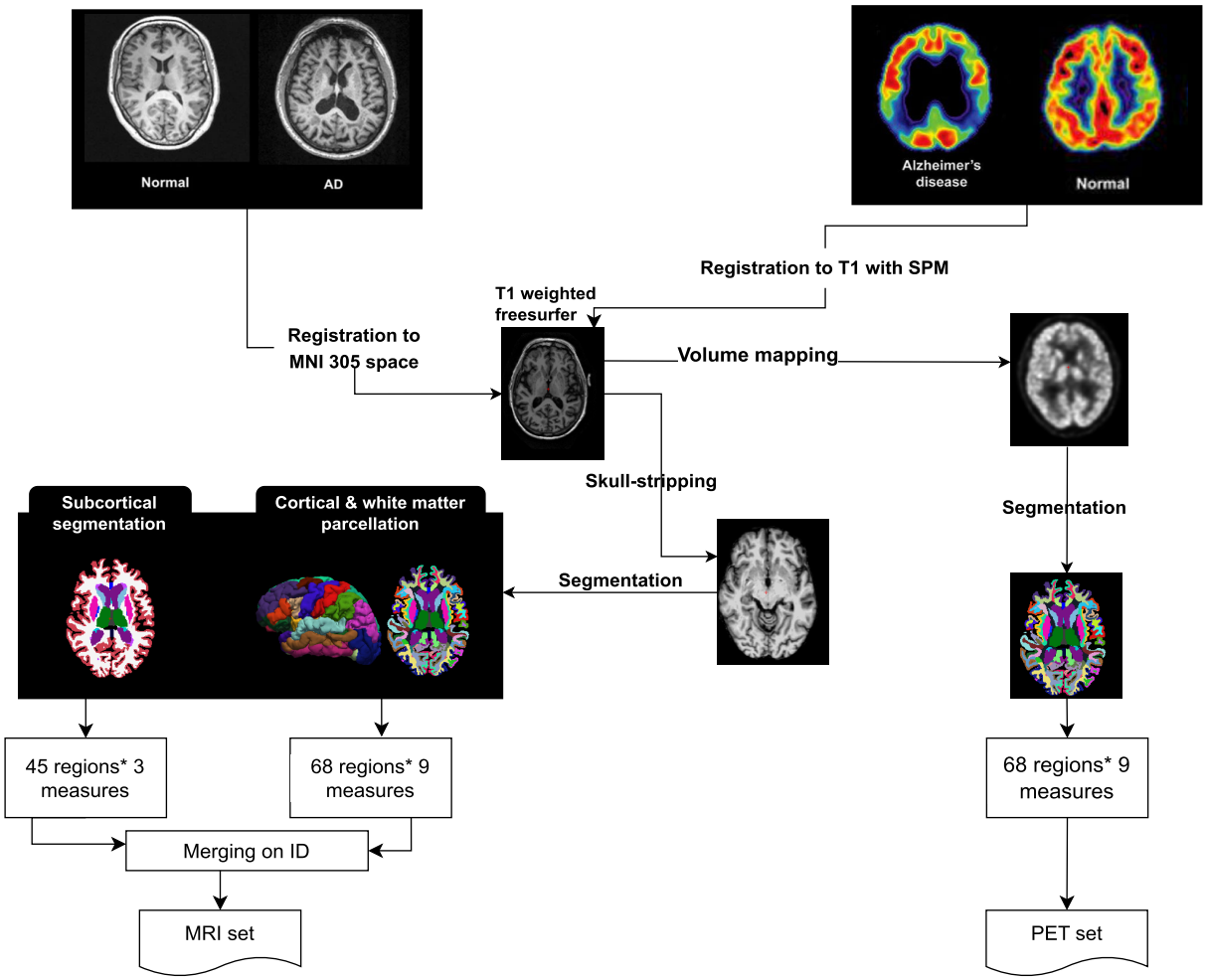


FIGURE 1. MRI and PET freesurfer processing pipeline.

To maintain the intra-relations, all subcortical and cortical regions from each modality are extracted and concatenated into the same training set. The selection method RC, used in feature selection considers all the features to identify the relevant ones, which further emphasizes the significance of intra-relations.

3) SAME-SUBJECT-MODALITIES-INTERACTIONS SSMI

The principle behind this criterion is to capture identical brain regions concurrently using different modalities or scan protocols. This approach offers the benefit of creating a feature set that captures the correlations and interactions between the corresponding regions in both PET and MRI modalities. Such a strategy can potentially enhance the performance of a classification model as it provides more information about each region by considering values for this region from both modalities. Additionally, employing the same segmentation for both modalities ensures the application of the same boundaries, enabling a more direct comparison between the regions. Moreover, the SSMI set not only provides additional features to improve the classification model's performance but also enriches the MRI and PET sets. This addition of new information can improve the training set's performance for classification models in AD classification tasks. To extract the SSMI.

We followed a three-step process illustrated in Figure 3 and explained in the following sections.

a: STEP1: CONSISTENT SEGMENTATION OF MRI AND PET

To ensure consistency between the boundaries of specific regions in PET and MRI, we applied the same segmentation to both modalities. This step enables a direct comparison of regions between the two modalities, with the region names serving as features for constructing the SSMI. Figure 4 illustrates the consistent segmentation of MRI and PET.

b: STEP 2: CREATING FEATURES WITH MULTIPLE VALUES

To improve our understanding of a particular brain region, we create features with multiple values from both PET and MRI modalities. This is achieved by identifying common regions between the two modalities and having the same boundaries, as determined in step 1.

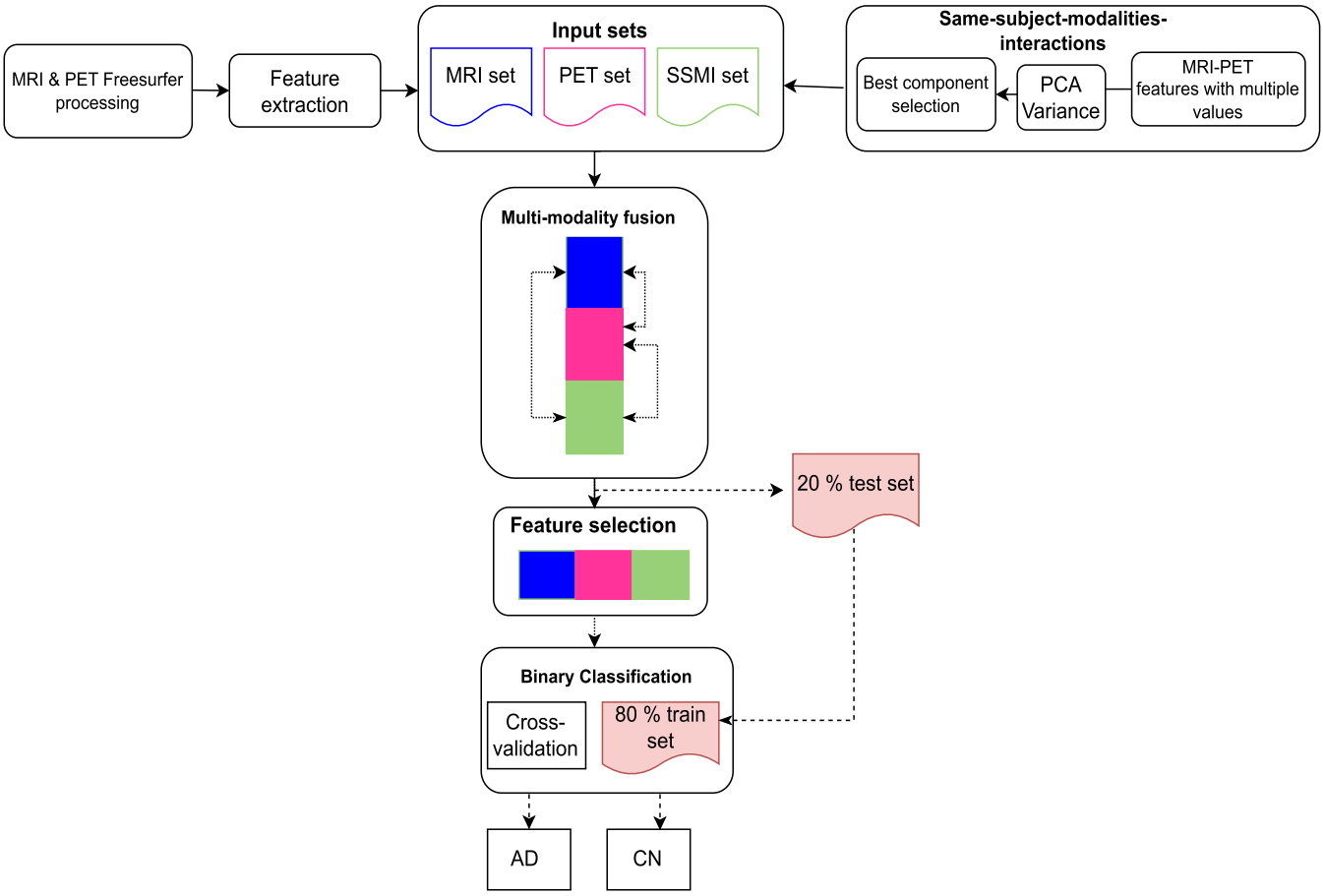


FIGURE 2. Workflow of the proposed architecture.

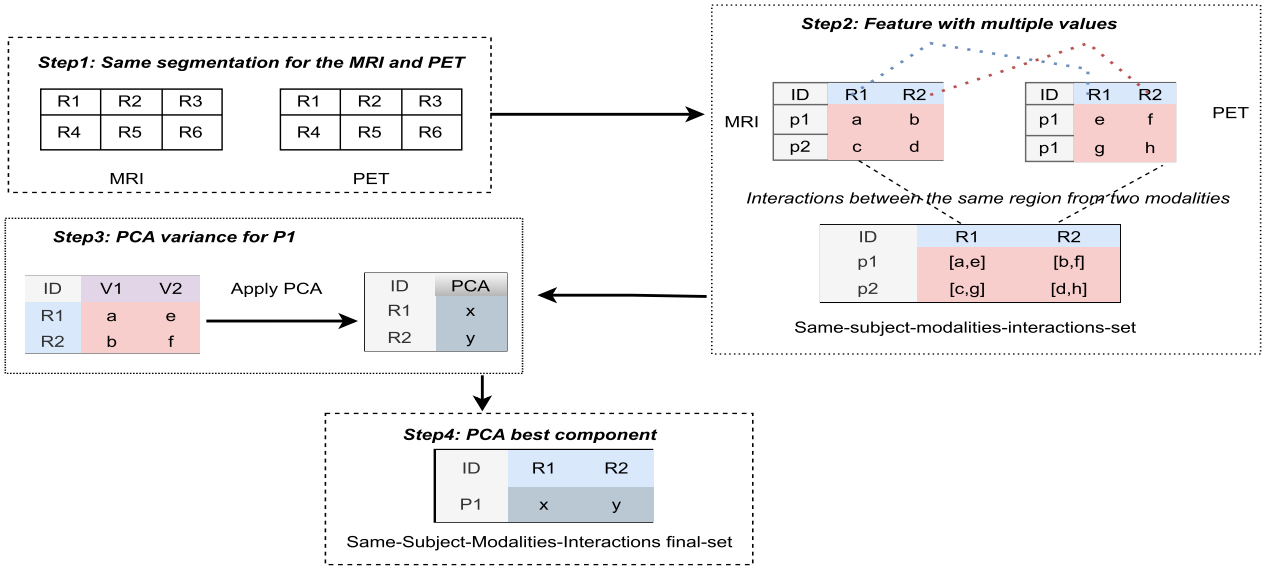


FIGURE 3. Extraction of "Same-Subject-Modalities-Interactions" set.

For instance, suppose that the MRI set contains three regions with three values, as shown in Table 2. The PET set also includes the same MRI regions but with different values, as presented in Table 3. We assume that a feature x referring to a region x will have the same raw MRI and PET values as shown in Table 4. By creating such features,

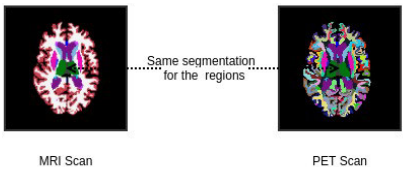


FIGURE 4. Same segmentation for MRI and PET regions.

TABLE 2. Example of MRI-set.

Id-patient	R_{MRI1}	R_{MRI2}	R_{MRI3}
Patient1	a	b	c
Patient2	d	e	f
Patient3	g	h	i

TABLE 3. Example of PET-set.

Id-patient	R_{PET1}	R_{PET2}	R_{PET3}
Patient1	j	k	l
Patient2	m	n	p
Patient3	q	s	v

TABLE 4. Example of "Same-Subject-Modalities-Interactions".

Id-patient	R1	R2	R3
Patient1	(a, j)	(b, k)	(c, l)
Patient2	(d, m)	(e, n)	(f, p)
Patient3	(g, q)	(h, s)	(i, v)

we can capture the correlations and interactions between the modalities for the same region and improve the performance of our classification model. We refer to this set of features as the SSMI set and present an example in Table 5.

In our research, we segmented MRI and PET into cortical and subcortical regions, resulting in two types of regions. The SSMI dataset contains 113 features, with 68 representing cortical regions and 45 representing subcortical regions.

For instance, let us examine the subcortical region called 'CC_Anterior'. The 'CC_Anterior' feature in the SSMI dataset comprises four values representing this region's measurements. These include the *NormDev*, *NormMean*, and *Volume* values for MRI modality, as well as the *SUV* value for PET modality. Let us assume that the feature 'CC_Anterior' has a value of (1,5,9,4).

As another example, consider the cortical region called 'bankssts'. The 'bankssts' feature in the SSMI dataset contains ten values representing this region's measurements. These include the *SurfArea*, *GrayVol*, *ThickAvg*, *ThickStd*, *MeanCurv*, *GausCurv*, *FoldInd*, *CurvInd*, *WM*, and *SUV* values for MRI and PET modalities, respectively. Let us assume that the feature 'bankssts' has a value of (12,20,30,41,12,5,8,25,15,2).

To ensure that the feature 'CC_Anterior' has the same number of values as the feature 'bankssts', we added zeros to the subcortical features. This way, both subcortical and cortical features have the same dimensions. Using the previous example, 'CC_Anterior' will have a value of (1,5,9,4,0,0,0,0,0,0). Table 5 provides an example of the SSMI set constructed from MRI and PET.

TABLE 5. Example of SSMI set used in the proposed work.

Id-patient	CC_Anterior	bankssts
Patient1	(1,5,9,4,0,0,0,0,0,0)	(12,20,30,41,12,5,8,25,15,2)

By including multiple values for the same region from both modalities, we can capture more information about that region and potentially improve the accuracy of our classification model.

c: STEP3: PCA VARIANCE

The SSMI generated in step 2 cannot be used directly in a classification task, as the model cannot train features with multiple values. To address this obstacle, we use a feature reduction technique to reduce the values of each feature to a single value. The reason why we use feature reduction and not feature selection is that we use feature selection on the 'bankssts' feature, the selected value will be one of the existing features, such as *SurfArea*. However, as we mentioned in Section III-A3.b, this individual feature already represents a feature in the MRI set. Therefore, using feature selection in this case would result in redundant features in the final dataset. This is why feature reduction is a better approach in this scenario, as it reduces the number of values in a feature to a single representative value, avoiding redundant features.

Principal Component Analysis (PCA) is a widely used technique for reducing the dimensionality of high-dimensional data [38]. PCA is a statistical technique used to reduce the dimensionality of high-dimensional datasets. It works by identifying the directions of maximum variance in the dataset and projecting the data onto these directions. The resulting new variables are called principal components, which are uncorrelated and can be used for further analysis. The main advantage of PCA over other techniques is that it allows for dimensionality reduction while preserving the most important information in the original data. This is achieved by identifying and retaining the principal components that explain the largest amount of variance in the data.

In our proposed architecture, we use PCA to represent a feature with the after creating the SSMI set based on the features with multiple values represented in step 2 of Section III-B3.b. We created a new input set for the PCA algorithm, where each row corresponds to a region and each column represents the values of this region as shown in Table 6. After applying PCA separately for each subcortical and cortical region, a separate eigenvector is obtained for each region. These new features are linear combinations of the original features and are orthogonal to each other. The eigenvectors are then concatenated horizontally to form a new feature set. Each row in this new feature set corresponds to a region, and each column represents the principal component that was extracted using PCA.

TABLE 6. Example PCA set.

Feature	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V_{10}
CC_Anterior	1	5	9	4	0	0	0	0	0	0
bankssts	12	20	30	41	12	5	8	25	15	2

It may make sense to apply PCA separately for each region to capture more specific information about each region's variance, particularly since there is no prior information on whether the regions are correlated or how an affected region with AD may influence the other ones.

d: STEP4: PCA BEST COMPONENT

Following PCA, we select the most informative principal component for each feature and reconstruct our data set so that each feature is represented by its optimal principal component, as described in Section III-B3.c. The resulting data set, final SSMI, consists of 113 features and each row corresponds to a patient. We will combine this final SSMI data set with the MRI and PET data sets to construct the final training set.

C. SETS-CONCATENATION

The fusion process involves combining the two modalities, MRI and PET, with the new set of SSMI. To construct the final training set for classification, the data matrix is concatenated. The total number of features is 973, consisting of 113 from SSMI, 45 from subcortical regions of MRI, 68 from cortical regions of MRI, and 113 from PET. To preserve the local structure information and ensure interactions between all subjects of the same or different modalities, appropriate preprocessing of the raw features is crucial before fusing all the data into the same matrix.

D. FEATURE SELECTION FOR INTER-RELATIONS

Feature selection is a critical process in machine learning that involves identifying the most informative subset of features from the original dataset while disregarding irrelevant, noisy, or redundant ones. The main objective of feature selection is to enhance the learning performance and reduce computational costs, leading to improved classification accuracy [39].

In the proposed architecture, we used feature selection to ensure interactions between all the features of all modalities. All the features are tested together in feature selection, proving that they interact. To perform feature selection, we used the Ridgeclassifier classifier, which is a parameter estimation technique applied to address the collinearity problem that frequently arises in multiple linear regression [40].

The RC is a linear model that can be used for binary classification and feature selection. It works by fitting a RC model to the data and selecting the features with the highest coefficients in the resulting model. RC is particularly useful for high-dimensional datasets as it solves a simple linear regression problem, making it fast and computationally efficient [41].

E. CLASSIFICATION

In our classification task, we aimed to predict whether patients have AD or are normal controls (CN). To accomplish this, we used five different classifiers: Logistic Regression (LR) [42], Linear Discriminant Analysis (LDA) [43], Support Vector Machine (SVM) [44], RC, and XGBoost [45].

IV. EXPERIMENTAL RESULTS

A. EXPERIMENTATION SCENARIOS

The experimentation scenarios involved several steps to validate the robustness of the proposed criteria:

- Firstly, single-modality classification tasks were performed on MRI, PET, and SSMI sets to demonstrate the validity of multimodality fusion. The prefix *N* was added to indicate experimentation without feature selection.
- Secondly, the MRI and PET sets were concatenated to assess the influence of intra-relations. The concatenated sets were referred to as 'MP.'
- Thirdly, the performance of the fused 'MPS' sets without feature selection was evaluated, and the prefix *N* was added to indicate experimentation without feature selection.
- Next, feature selection was then applied to 'MP' and 'MPS' sets to ensure the inter-relations. The prefix *S* was added to indicate experimentation with feature selection.
- Additionally, the proposed method was compared with a state-of-the-art method based on performance evaluation metrics and robust fusion criteria.
- Finally, the learning performance of the new dataset from ADNI-2 and ADNI-GO was assessed to evaluate the generalizability of the proposed method.

B. VALIDATION TECHNIQUES

We used two ML techniques in the experimentation to ensure prediction stability: cross-validation and train-test-split. Additionally, to assess the performance of the proposed framework on real data, we used another dataset, ADNI-2/ADNI-GO. For performance evaluation, we used different measures: accuracy, recall, precision, and F1-score.

1) CROSS-VALIDATION

Cross-validation is a technique used to assess the predictive performance of a model. It involves dividing the dataset into several subsets, known as folds. The model is trained on a subset of the data and tested on the remaining data, and this process is repeated for each fold. This technique helps to prevent overfitting and assesses the generalization ability of the model [46]. Cross-validation takes time because the training algorithm must be rerun k times. An alternative solution is train-test-split.

2) TRAIN-TEST-SPLIT

Train-test split is a frequently employed method in ML for evaluating the performance of a prediction. It entails separating the data into a training set and a test set, with

TABLE 7. Demographic information of the subjects considered in the validation set.

Classes	AD	CN
Number of subjects	16	65
Female/Male	2/14	10/55
Age(year)AD	74.93 ±7.21	75.81 ±6.25

the model being trained on the training set and then evaluated on the test set. The objective is to assess how well the model performs on fresh, unobserved data. In our study, we used an 80/20 split for training and testing, respectively. However, previous researchers have encountered a methodological issue when performing feature selection. If feature selection is performed after dividing the data into training and test sets, the selected features may already be known in the test set during prediction evaluation, leading to biased results. Therefore, in our study, we performed feature selection only on the training set to make predictions, and then applied the selected features to evaluate the prediction performance on the test set. This ensures that the model is evaluated on an independent dataset and can produce reliable results.

3) VALIDATION DATASET ADNI-2/GO

ADNI-2/ADNI-GO is another dataset used to evaluate the performance of the proposed framework on real data. This dataset is a collection of MRI, PET, and clinical data from Alzheimer's patients and normal controls. It allows for the evaluation of the proposed method's generalizability to different datasets. Table 7 presents information about the participants. We followed the same preprocessing steps used for the ADNI set to prepare and process the subjects in the validation set.

C. PERFORMANCE EVALUATION METRICS

Accuracy, Precision, Recall, and F1-score are the different metrics employed to assess the efficacy of a classifier [47]. When evaluating the performance of a model on an unbalanced dataset, accuracy alone may not be a reliable metric. This is due to the fact that a classifier that merely predicts the majority class for every instance in the dataset can achieve high accuracy, but such a model may not be applicable in real-world settings in which we wish to discover the minority class. Many evaluation metrics, including precision, specificity, recall, F1 score, and AUC, are utilized to address this issue [48].

1) ACCURACY

A term to refer the proportion of correctly identified points relative to the total number of points. The main disadvantage of this metric is that it needs to evaluate the performance of imbalanced data reliably.

2) PRECISION

Precision is the fraction of the correctly classified instances from the total classified instances.

3) SPECIFICITY

Specificity is defined as the proportion of actual negative cases that are correctly identified by the model, also known as true negatives (TN), over the total number of actual negative cases.

4) RECALL

Recall is the proportion of instances properly classified out of all instances classified.

5) F1-SCORE

F1-score is the harmonic mean of precision and recall.

6) AUC

AUC measures the ability of a model to distinguish between positive and negative classes by calculating the area under the Receiver Operating Characteristic (ROC) curve.

D. EXPERIMENTAL RESULTS OF SINGLE-MODALITY CLASSIFICATION TASK

The study evaluated the classification performance of three modalities, MRI, PET, and SSMI, separately using train-test-split and cross-validation methods, and compared the results with the fused set. Table 8 and Table 9 show the classification results of each modality using train-test-split and cross-validation, respectively.

The MRI modality had the highest accuracy among all modalities, achieving an accuracy of 91.89% using Xgboost and LDA in train-test-split. The PET set had the lowest accuracy, with an accuracy of 67.56% using SVM in train-test-split and 64.06% by RC in cross-validation. Similarly, the SSMI set's best result was achieved using Xgboost, with an accuracy of 62.16% in train-test-split and 62.28% by SVM in cross-validation.

In terms of precision, the MRI modality showed the highest results among all modalities, with a precision of 100% using LDA in train-test-split and 90.90% using Xgboost in train-test-split. Specificity was also highest for MRI, with a specificity of 100% using LDA in train-test-split. The PET set had the lowest precision among all modalities, with a precision of 11.11% using LR in train-test-split. The SSMI set had the lowest specificity, with a specificity of 52.38% using LDA in cross-validation.

E. INFLUENCE OF DIFFERENT RELATIONS

In this section, we investigate the impact of different relations on the classification performance. Specifically, we examine the intra-relations between subjects of the same modality and inter-relations between different modalities.

1) INTRA-RELATIONS

To evaluate the performance of our models in predicting intra-relations between MRI and PET modalities, we concatenated the two sets without feature selection, denoted as *N*MP. The classification performance of the models using

TABLE 8. Classification performances (%) of single modality using train-test-split.

Modality	Classif	Accy	Pre	Spec	Recall	F1	AUC
NMRI	LR	86.48	88.88	96.00	66.66	76.19	81.33
	LDA	91.89	100.0	100.0	75.00	85.71	87.51
	SVM	81.08	85.71	96.00	50.00	63.15	73.01
	RC	72.97	60.00	84.01	50.01	54.54	67.10
	Xgboost	91.89	90.90	96.00	83.33	86.95	89.66
NPET	LR	51.35	11.11	69.23	09.09	10.00	39.16
	LDA	51.35	26.66	57.69	36.36	30.76	47.02
	SVM	67.56	00.00	96.15	00.00	00.00	48.07
	RC	54.05	12.50	10.52	09.09	73.07	41.08
	Xgboost	56.75	22.22	73.07	18.18	20.00	45.62
NSSMI	LR	48.64	00.00	85.71	00.00	00.00	42.85
	LDA	45.94	37.50	52.38	37.50	37.50	44.94
	SVM	48.64	00.00	85.71	00.00	00.00	42.85
	RC	48.64	00.00	85.71	00.00	00.00	42.85
	Xgboost	62.16	60.00	80.95	37.50	46.15	59.22

TABLE 9. Classification performances (%) of single modality using cross-validation.

Modality	Classif	Accy	Pre	Spec	Recall	F1	AUC
NMRI	LR	85.78	86.44	92.98	73.91	79.68	83.44
	LDA	81.92	78.12	87.71	72.46	75.18	80.09
	SVM	84.67	85.96	92.98	71.01	77.77	81.99
	RC	78.12	74.57	86.84	63.76	68.75	75.30
	Xgboost	84.70	81.53	89.47	76.81	79.10	83.14
NPET	LR	61.34	42.85	92.98	08.69	14.45	50.83
	LDA	55.23	38.98	68.42	33.33	35.93	50.87
	SVM	62.45	00.00	100.0	00.00	00.00	50.00
	RC	64.06	57.14	92.10	17.39	26.66	54.74
	Xgboost	54.26	35.84	70.10	27.53	31.14	48.85
NSSMI	LR	61.75	40.00	97.36	02.89	05.40	50.13
	LDA	43.01	32.32	41.22	46.37	38.09	43.80
	SVM	62.28	00.00	100.0	00.00	00.00	50.00
	RC	60.67	43.47	88.59	14.49	21.73	51.54
	Xgboost	52.39	35.93	64.03	33.33	34.58	48.68

train-test-split and cross-validation is shown in Table 10 and Table 11, respectively.

In train-test-split experimentation, LR, LDA, and Xgboost achieved the highest accuracy rates of 89.18%, 86.48%, and 86.48%, respectively, in the intra-relation classification task using train-test split. However, LDA and SVM had a lower recall rate of 63.63% compared to the other classifiers. LDA and Xgboost, achieved the best specificty of 96.15%. RC had the lowest accuracy rate of 75.67% and the lowest recall rate of 54.54%.

In cross-validation experimentation, the accuracy of the classification models ranged from 76.60% to 86.28%, with Xgboost having the highest accuracy. LR had the highest precision and specificity among all the classifiers, while Xgboost had the highest recall. RC had the lowest accuracy, precision, specificity, and recall. The F1 score and AUC ranged from 68.75% to 81.75% and 74.84% to 85.31%, respectively, with Xgboost having the highest F1 score and AUC.

2) SAME-SUBJECT-MODALITIES-INTERACTIONS

This section evaluates the performance of the fused sets (MRI, PET, and SSMI) with a specific focus on the impact of SSMI. The fused set is denoted as *N*MPS. Table 12 shows

TABLE 10. Classification performances (%) of intra-relation with train-test-split.

Modality	Classif	Accy	Pre	Spec	Recall	F1	AUC
NMP	LR	89.18	81.81	92.30	81.81	81.81	87.06
	LDA	86.48	87.5	96.15	63.63	73.68	79.89
	SVM	83.78	77.77	92.30	63.63	70.00	77.97
	RC	75.67	60.00	84.61	54.54	57.14	69.58
	Xgboost	86.48	87.5	96.15	63.63	73.68	79.89

TABLE 11. Classification performances (%) of intra-relation with cross-validation.

Modality	Classif	Accy	Pre	Spec	Recall	F1	AUC
NMP	LR	85.78	86.44	92.98	73.91	79.68	83.44
	LDA	82.98	79.68	88.59	73.91	76.69	81.25
	SVM	83.01	83.92	92.10	68.11	75.20	80.11
	RC	76.60	69.11	81.57	68.11	68.61	74.84
	Xgboost	86.28	82.35	89.47	81.15	81.75	85.31

TABLE 12. Classification performances (%) of *N*MPS set using train-test-split.

Modality	Classif	Accy	Pre	Spec	Recall	F1	AUC
NMPS	LR	91.89	83.33	92.30	90.90	86.95	91.60
	LDA	86.48	87.50	96.15	63.63	73.68	79.89
	SVM	86.48	63.63	96.15	73.68	79.89	79.89
	RC	78.37	66.66	88.46	54.54	60.00	71.50
	Xgboost	86.48	87.50	96.15	63.63	73.68	79.89

TABLE 13. Classification performances (%) for *N*MPS set using cross-validation.

Modality	Classif	Accy	Pre	Spec	Recall	F1	AUC
NMPS	LR	85.17	85.00	92.10	73.91	79.06	83.00
	LDA	84.06	82.25	90.35	73.91	77.86	82.13
	SVM	81.90	84.61	92.98	63.76	72.72	78.37
	RC	77.69	71.21	83.33	68.11	69.62	75.72
	Xgboost	86.81	84.61	91.22	79.71	82.08	85.46

the results of the classification task using train-test-split, and Table 13 shows the results using cross-validation.

In train-test-split experimentation, LR achieved the highest accuracy rate of 91.89% and a recall rate of 90.00%. LDA and Xgboost showed the highest precision rate of 87.50%. LDA achieved an accuracy rate of 86.48%, while RC achieved a slightly lower accuracy rate of 78.37%. However, SVM showed the highest specificity rate of 96.15% next to LDA and Xgboost. LR achieved the best performance in terms of F1 and AUC among the classifiers on the *N*MPS set.

In cross-validation experimentation, the accuracy of the classifiers ranged from 77.69% to 86.81%, with Xgboost having the highest accuracy. LR had the highest precision and specificity among all the classifiers, while SVM had the highest specificity. RC had the lowest accuracy, precision, and specificity. The F1 score and AUC ranged from 69.62% to 82.08% and 75.72% to 85.46%, respectively, with Xgboost having the highest F1 score and AUC. The results obtained using cross-validation are consistent with those obtained using train-test-split, with some variations in the performance of each classifier.

3) INTER-RELATIONS

Table 14 and Table 15 show the results of the train-test-split and cross-validation experiments, respectively, where feature selection was applied to ensure inter-relations for each

TABLE 14. Classification performances (%) for SMP using train-test-split.

Modality	Classif	Accy	Pre	Spec	Recall	F1	AUC
SMP	LR	86.48	85.71	66.60	96.06	90.56	81.33
	LDA	67.56	78.26	58.33	72.00	74.99	65.16
	SVM	86.48	85.71	66.66	96.00	90.56	81.33
	RC	86.48	88.46	75.00	92.00	90.19	83.50
	Xgboost	86.48	91.66	83.33	88.00	89.79	85.66

TABLE 15. Classification performances (%) for SMP using cross-validation.

Modality	Classif	Accy	Pre	Spec	Recall	F1	AUC
SMP	LR	93.39	94.73	91.30	94.73	94.73	93.02
	LDA	96.66	98.21	97.10	96.49	97.34	96.79
	SVM	94.50	96.42	94.20	94.73	95.57	94.46
	RC	96.66	96.55	94.20	98.24	97.39	96.22
	Xgboost	85.29	87.17	78.26	89.47	88.31	83.86

modality. For the MP set, 52 features were selected, which was equivalent to the number of features in the MPS.

Based on Table 14, it can be seen that LR, SVM, RC, and Xgboost achieve relatively high accuracy, ranging from 86.48% to 86.81%. LDA, on the other hand, has a relatively low accuracy of 67.56%. The highest precision is achieved by Xgboost with 91.66%, while the highest recall is obtained by LR with 96.06%. The AUC values range from 65.16% to 85.66%.

According to Table 15, it can be observed that LR, LDA, SVM, RC, and Xgboost achieved relatively high accuracy, ranging from 85.29% to 96.66%. Xgboost had the lowest accuracy, while LDA had the highest accuracy. The highest precision and specificity are achieved by LDA with 98.21% and 97.10%, respectively, while the highest recall is obtained by RC with 98.24%. The AUC values range from 83.86% to 96.79%, where LDA has the highest AUC value, while Xgboost has the lowest.

Table 16 and Table 17 present the classification performance results of SMPS using feature selection in train-test-split and cross-validation experiments, respectively. The optimal number of selected features was found to be 59, and Figure 5 and Figure 6 show the corresponding *S*MPS ROC curves.

In train-test-split experiment, LR, SVM, and RC achieved the highest accuracy score with 97.29%. LDA had a lower accuracy score with 94.59%, while Xgboost had the lowest score with 91.89%. Regarding precision, RC had the highest score with 100%, followed by LR and SVM with 91.66%. Xgboost had the lowest score with 83.33%. For specificity, RC had the highest score with 100%, followed by LR, LDA, and SVM with 96.15%. Xgboost had the lowest score with 92.30%. In terms of recall, LR and SVM achieved a perfect score of 100%, while LDA, Xgboost, and RC had a score of 90.90%. For F1-score, LR and SVM had the highest score with 95.65%, followed by RC with 95.23%. Xgboost had the lowest score with 86.95%. Finally, in terms of AUC score, LR and SVM had the highest score with 98.07%, followed

TABLE 16. Classification performances (%) for SMPS using Train-test-split.

Modality	Classif	Accy	Pre	Spec	Recall	F1	AUC
SMPS	LR	97.29	91.66	96.15	100.0	95.65	98.07
	LDA	94.59	90.90	96.15	90.90	90.90	93.53
	SVM	97.29	91.66	96.15	100.0	95.65	98.07
	RC	97.29	100.0	100.0	90.90	95.23	95.45
	Xgboost	91.89	83.33	92.30	90.90	86.95	91.60

TABLE 17. Classification performances (%) for SMPS using cross-validation.

Modality	Classif	Accy	Pre	Spec	Recall	F1	AUC
SMPS	LR	93.97	94.78	91.30	95.61	95.19	93.45
	LDA	97.33	97.39	95.65	98.24	97.81	96.94
	SVM	94.56	96.42	94.20	94.73	95.57	94.46
	RC	98.94	98.27	97.10	100.0	99.13	98.55
	Xgboost	88.50	91.15	85.50	90.35	90.74	87.92

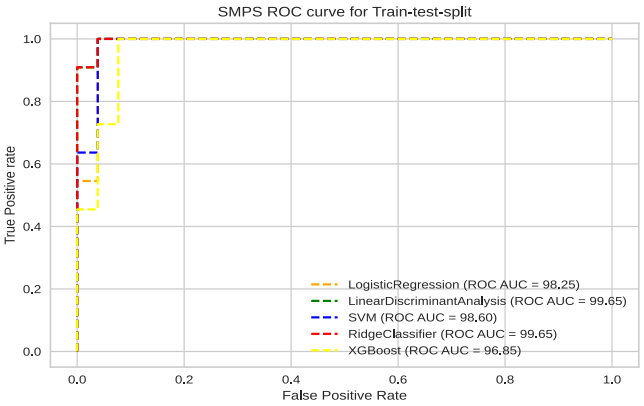


FIGURE 5. SMPS ROC curve for each model in Train-test-split.

by RC with 95.45%, while Xgboost had the lowest score with 91.60%.

In the cross-validation experiment, the highest accuracy score of 98.94% was achieved by RC, while the lowest score of 88.50% was achieved by Xgboost. Similarly, the highest precision score of 98.27% was also achieved by RC, while the lowest score of 91.15% was achieved by Xgboost. The highest specificity score of 97.10% was achieved by RC, while the lowest score of 85.50% was achieved by Xgboost. RC achieved the highest recall score of 100.0%, while SVM achieved the lowest score of 94.73%. Additionally, the highest F1 score of 99.13% was achieved by RC, while Xgboost achieved the lowest score of 90.74%. Finally, the highest AUC score of 98.55% was achieved by RC, while Xgboost achieved the lowest score of 87.92%. Overall, these results indicate that RC and LDA outperformed the other models across most of the metrics, while Xgboost consistently had the lowest scores.

F. COMPARISON WITH OTHER EXISTING METHODS

The proposed approaches are compared to state-of-the-art methods based on both robust fusion criteria and accuracy in Table 18. It is worth noting that previous publications

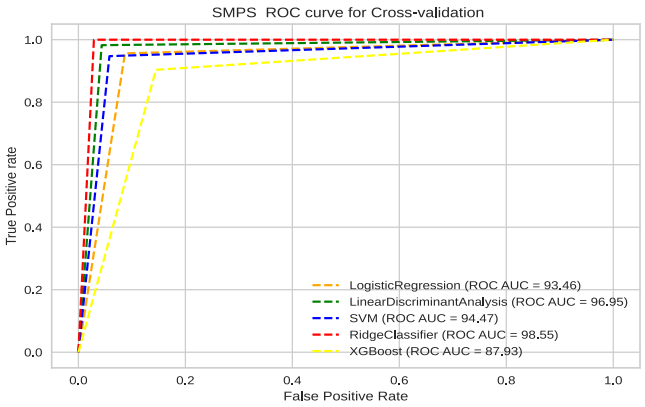


FIGURE 6. SMPS ROC curve for each model in cross-validation.

TABLE 18. Comparison with other existing methods.

Ref	Inter	Intra	SSMI	Structure information	Acc	AUC
Early						
[20]	+	-	-	+	94.70	98.00
[49]	+	-	-	-	90.70	94.50
[18]	+	-	-	-	88.80	96.10
[19]	+	-	-	-	85.70	-
Intermediate						
[50]	+	+	-	+	93.30	97.00
[24]	-	+	-	+	94.37	97.24
[8]	+	+	-	+	96.76	97.03
[17]	+	+	-	+	97.20	98.00
[51]	+	+	-	+	94.58	-
Late						
[26]	-	-	-	+	93.00	-
[27]	-	-	-	+	94.80	98.6
[25]	-	-	-	+	97.00	97.00
[28]	-	-	-	+	94.80	-
Proposed method						
	+	+	+	+	98.94	98.55

did not incorporate the proposed robust fusion criteria into their frameworks. In the early fusion approach, inter-relations were guaranteed, and in the intermediate fusion approach, inter-intra relations were imposed while preserving structure information. However, this was not ensured in the late fusion approach, and none of the relations were guaranteed. Our work achieved higher accuracy than state-of-the-art methods. The closest accuracy to the proposed framework was 97.20%.

G. SELECTED BIOMARKERS

Table 19 shows the features selected by RC, where the region and its measures are listed, and the column indicates the initial data source. The segmentation identifies the region segmentation. We summarize the selected features in the following four points:

1) MRI FEATURES

MRI features were selected from cortical, subcortical, and white matter segmentation.

a: RIGHT CORTEX

The **parahippocampal** region was selected with three measures: *WM − volume*, *SurfArea*, and *ThickAvg*. The **superior frontal** region has two measures selected: *WM − volume*, and *MeanCurv*. The **paracentral** region has two measures selected: *ThickStd* and *GrayVol*. Additionally, the **entorhinal** region has been selected with *ThickAvg*, and *GrayVol*.

b: LEFT CORTEX

The **frontalpole** region was selected with two measures: *ThickStd*, and *MeanCurv*. The **paracentral** region was also selected with *FoldInd* and *GrayVol*. The **inferior temporal** region was selected with *ThikAvg* and *GrayVol*.

c: SUBCORTICAL

The **Left-Hippocampus** was selected with three measures: *normDev*, *NormMean*, and *Volume*. Additionally, the **Right-Hippocampus** was selected with *NormMean* and *Volume*.

2) SSMI FEATURES

From SSMI, only *non − WM − hypointensities* and *ctx − lh − isthmuscingulate* were selected.

3) PET FEATURES

The features of PET were not selected as the best set of features.

H. EXPERIMENTAL RESULTS OF VALIDATION SET

The performance of our proposed method shows promising results. However, it is crucial to evaluate the method on an additional validation dataset to ensure its generalizability. We used only the 59 features selected by RC as input and evaluated the performance using both train-test-split and cross-validation on the validation set.

1) VALIDATION SET RAW-DATA

Table 20 and Table 21 present the results of train-test split and cross-validation for all dataset features (original raw data), respectively. We denote it as VDR.

In train test split, it can be observed that the highest accuracy achieved is 88.23% by RC, while the lowest accuracy is 70.58% obtained by LDA, SVM, and Xgboost. LR, LDA, SVM, RC, and Xgboost achieved different levels of performance in terms of precision, recall, and AUC. LR, RC obtained the highest precision of 100.0%, while RC achieved the highest recall of 60.0%. The AUC values ranged from 50.00% to 80.00%, with RC having the highest AUC value of 80.00%, while SVM had the lowest with 50.00%.

TABLE 19. List of regions selected y RC.

Region	Measure	Set	Segmentation
Caudal anterior cingulate	CurvInd	MRI	Right-cortex
Rostral anterior cingulate	CurvInd	MRI	Right-cortex
Frontalpole	CurvInd	MRI	Right-cortex
Parahippocampal	WM,SurfArea , ThickAvg	MRI	Right-cortex
Superior frontal	WM, MeanCurv	MRI	Right-cortex
Bankssts	ThickStd	MRI	Right-cortex
Fusiform	ThickStd	MRI	Right-cortex
Lateral occipital	ThickStd	MRI	Right-cortex
Lateral orbitofrontal	ThickStd	MRI	Right-cortex
Paracentral	ThickStd, GrayVol	MRI	Right-cortex
Temporal pole	ThickStd	MRI	Right-cortex
Insula	GausCurv	MRI	Right-cortex
Entorhinal	ThickAvg, GrayVol	MRI	Right-cortex
Precuneus	ThickAvg	MRI	Right-cortex
Superior parietal	MeanCurv	MRI	Right-cortex
Inferior parietal	GrayVol	MRI	Right-cortex
Pars triangularis	GrayVol	MRI	Right-cortex
Transverse temporal	GrayVol	MRI	Right-cortex
Superior temporal	ThickStd	MRI	Left-cortex
Frontalpole	ThickStd , MeanCurv	MRI	Left-cortex
Insula	ThickStd	MRI	Left-cortex
Medial orbitofrontal	FoldInd	MRI	Left-cortex
Paracentral	FoldInd, GrayVol	MRI	Left-cortex
Superior parietal	FoldInd	MRI	Left-cortex
Entorhinal	ThickAvg	MRI	Left-cortex
Inferior temporal	ThickAvg, GrayVol	MRI	Left-cortex
Pars opercularis	ThickAvg	MRI	Left-cortex
Pericalcarine	ThickAvg	MRI	Left-cortex
Posterior cingulate	ThickAvg	MRI	Left-cortex
Rostral middle frontal	ThickAvg	MRI	Left-cortex
tTemporal pole	ThickAvg	MRI	Left-cortex
Lateral orbito frontal	MeanCurv	MRI	Left-cortex
Superior frontal	MeanCurv	MRI	Left-cortex
Supramarginal	MeanCurv	MRI	Left-cortex
Left-Cerebellum-Cortex	normDev	MRI	Subcortical
Left-Hippocampus	normDev, NormMean, Volume	MRI	Subcortical
Right-Caudate	normDev	MRI	Subcortical
Left-Inf-Lat-Vent	NormMean	MRI	Subcortical
Left-Caudate	NormMean	MRI	Subcortical
Right-Hippocampus	NormMean, Volume	MRI	Subcortical
Right-Amygdala	NormMean	MRI	Subcortical
Non-WM-hypointensities	NormMean	MRI	Subcortical
CC-Mid-Posterior	NormMean	MRI	Subcortical
Brain-Stem	Volume	MRI	Subcortical
Left-Amygdala	Volume	MRI	Subcortical
Right-Putamen	Volume	MRI	Subcortical
Right-Accumbens-area	Volume	MRI	Subcortical
Non-WM-hypointensities	PCA-variance	SSMI	Subcortical
Ctx-lh-isthmuscingulate	PCA-variance	SSMI	left-cortex

TABLE 20. Classification performances (%) for VDR using train-test-split.

Modality	Classif	Accy	Pre	Spec	Recall	F1	AUC
VDR	LR	76.47	100.0	100.0	20.00	33.33	60.00
	LDA	70.58	50.00	91.66	20.00	28.57	55.83
	SVM	70.58	00.00	100.0	00.00	00.00	50.00
	RC	88.23	100.0	100.0	60.00	74.99	80.00
	Xgboost	70.58	50.0	91.66	20.00	28.57	55.83

The results of cross-validation indicate that LR, LDA, RC, and Xgboost achieved accuracy ranging from 83.75% to 86.25%, which is relatively high. However, SVM showed poor performance with 80.0% accuracy. The precision values ranged from 00.00% to 75.00%, where LR achieved the highest precision, and SVM achieved the lowest. The recall values were between 00.00% and 50.00%, where RC achieved the highest recall, and SVM had the lowest. The AUC values ranged from 50.00% to 72.69%, where RC had the highest AUC value, and SVM had the lowest.

2) SELECTED FEATURES

Table 22 and Table 23 represent the results of train-test-split and cross-validation, respectively, using only the 59 features by RC in Section IV-E3. We denoted this with *S*VDR.

TABLE 21. Classification performances (%) for VDR using cross-validation.

Modality	Classif	Accy	Pre	Spec	Recall	F1	AUC
VDR	LR	85.00	75.0	96.92	37.50	50.00	67.21
	LDA	83.75	66.66	95.38	37.5	48.00	66.44
	SVM	80.00	00.00	100.0	00.00	00.00	50.00
	RC	86.25	72.72	95.38	50.00	59.25	72.69
	Xgboost	83.75	66.66	95.38	37.5	48.00	66.44

TABLE 22. Classification performances (%) for *S*VDR using train-test-split.

Modality	Classif	Accy	Pre	Spec	Recall	F1	AUC
SVDR	LR	94.11	100.0	100.0	80.0	88.88	90.0
	LDA	82.35	62.5	75.0	100.0	76.92	87.5
	SVM	88.23	100.0	100.0	60.0	74.99	80.0
	RC	100.0	100.0	100.0	100.0	100.0	100.0
	Xgboost	94.11	100.0	100.0	80.0	88.88	90.0

TABLE 23. Classification performances (%) for *S*VDR using cross-validation.

Modality	Classif	Accy	Pre	Spec	Recall	F1	AUC
SVDR	LR	92.50	92.75	68.75	98.46	95.52	83.60
	LDA	95.00	96.92	87.50	96.92	96.92	92.21
	SVM	92.50	92.75	68.75	98.46	95.52	83.60
	RC	98.75	98.48	93.75	100.0	99.23	96.80
	Xgboost	86.25	88.57	50.0	95.38	91.85	72.69

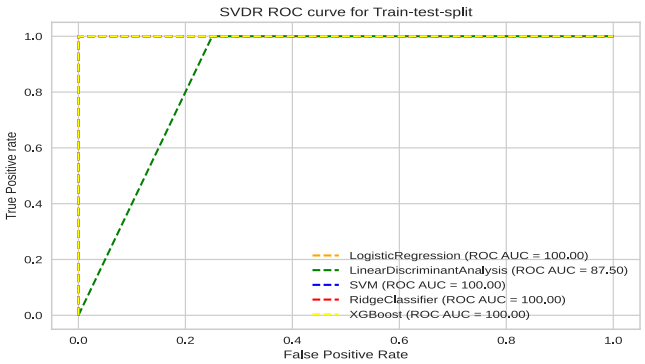


FIGURE 7. SVDR ROC curve for each model in train-test-split.

Figure 7, Figure 8 plot the *S*VDR ROC curve for each model in train-test-split and cross-validation, respectively.

In train-test-split experimentation, all the classifiers achieved high accuracy ranging from 88.23% to 100%, except LDA, which showed poor performance with 82.35% accuracy. LR achieved the highest precision with 100.0%, while RC achieved the highest recall with 100.0%. RC also had the highest AUC value of 100.0%.

In cross-validationt experimentation, all the classifiers except Xgboost achieved accuracy ranging from 92.50% to 98.75%, with RC achieving the highest accuracy. RC also achieved the highest precision, specificity, recall, and AUC value. Xgboost showed the lowest performance with 86.25% accuracy and the lowest AUC value.

V. DISCUSSION

In our study, we developed a machine learning framework that fuses MRI and PET data to improve the accuracy of

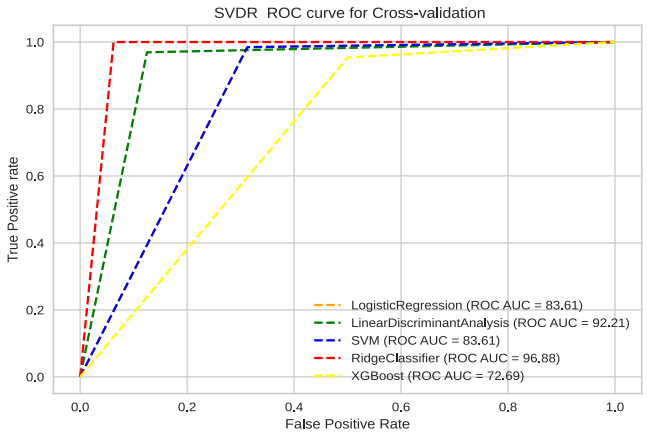


FIGURE 8. SVDR-cross ROC curve for each model in cross-validation.

AD diagnosis. To ensure robust performance and extract complementary information, we proposed several criteria for constructing the fusion framework. Our approach was assessed on a dataset of 183 ADNI subjects and evaluated using both train-test-split and cross-validation methods. We also compared our method to state-of-the-art approaches in the field. Based on our results, we can summarize the following findings: Firstly, we evaluated the prediction performance using only one modality. The MRI modality was found to have the highest accuracy using LDA and LR with 91.89% and 85.78% accuracy in train-test-split and cross-validation, respectively. The good results of MRI set would be form many reasons. Firstly, MRI can provide high-resolution images of the brain that can capture structural changes associated with neurodegeneration, such as brain atrophy or changes in gray matter density [52]. Moreover, MRI techniques can also capture changes in functional connectivity between different brain regions, providing additional information about the brain's network architecture [53]. Also, the MRI dataset downloaded from ADNI was preprocessed, ensuring its quality.

PET provided poor prediction performance with both validation techniques. Although PET has a high selectivity for tau $A\beta$ deposits, which are a symptom of AD [54], the existence of amyloid-positive PET only reflects the existence of amyloid neuropathology in vivo but does not confirm the existence of AD. Additionally, cognitively normal subjects (even young and healthy volunteers) have tested positive for amyloid-positive PET. This means that amyloid-positive PET is not a definitive diagnostic criterion for AD [55]. Moreover, some researchers have debated excluding AD only based on a negative amyloid scan [56]. Therefore, these limitations of PET may explain its poor performance predicting AD in this study.

The SSMI classification task is a crucial aspect of this study, as it involved using both MRI and PET values. However, despite the combination of both modalities, poor prediction accuracy was obtained. SSMI, or the specific set of multimodality imaging data used in this study, included both

MRI and PET values. The excellent learning performance achieved using MRI indicates that the poor prediction accuracy of SSMI can be attributed to the limitations of PET.

The NMP fusion approach resulted in a significant improvement in classification accuracy, with an increase from 86.48% to 89.18% for LR in the train-test-split, representing a 4% improvement. The recall rate also showed improvement from 66.66% to 81.81%, and the F1-score increased from 76.19% to 81.81%. The AUC value also improved from 81.33% to 87.06%. In addition, the NMP fusion approach improved the LDA accuracy in cross-validation from 81.92% to 82.98%. The accuracy of Xgboost also increased from 84.70% to 86.28%, and its AUC improved from 83.14% to 85.31%. These findings demonstrate the significant benefit of combining the MRI and PET modalities through fusion techniques to improve learning performance. In the imbalanced dataset, the *N*MP fusion showed a significant improvement in term of recall, F1-score, and AUC. This indicates that the fusion approach was effective in addressing the challenge of imbalanced data by improving the model's ability to detect positive cases. In an imbalanced dataset, accuracy alone can be a misleading metric as it can be high, simply due to the majority class being predicted correctly. In this scenario, metrics such as recall, F1-score, and AUC become more important as they take into account the performance of the classifier on the minority class [57].

The accuracy of LDA and Xgboost decreased from 91.89 to 86.48, which could be attributed to the poor quality of PET scans. PET typically has lower resolution than MRI and may require enhancement to improve its informativeness. In our study, we used PET scans that were downloaded without any preprocessing and processed them using Freesurfer. While the MRI set showed good performance, indicating the robustness of Freesurfer as a tool for MRI processing, PET processing may still require additional enhancements. Furthermore, PET was segmented based on MRI segmentation, meaning that any errors in the MRI segmentation process could have affected PET processing and ultimately impacted classification accuracy.

The results of NMPS experiment, indicate a significant improvement in the classification performance. The highest accuracy in MPS for train-test-split increased from 89.18% in MP to 91.89%, while the highest recall increased from 81.81% to 90.90%, and the highest recall rate increased to 86.95% from 81.81%. The F1-score also increased form 86.95% to 90.56%. Moreover, the highest AUC increased from 87.06 to 91.60. While the scores remain stable in cross-validation. These findings confirm the positive influence of the SSMI set on classification task.

The results of SMP experiment indicate a significant improvement in classification performance when RC selection was applied to MP. The highest precision in MP for train-test-split increased from 87.5% to 91.66%, while the highest recall increased from 81.81% to 96.06%, and the highest F1-score increased to 90.06% from 81.81%. Moreover, the highest AUC increased from 87.06% to 85.66%.

In addition, the highest accuracy in NMP for cross-validation increased from 86.81% to 96.66%, while the highest precision increased from 85.00% to 98.21%, and the highest specificity increased to 97.10% from 92.98%. Furthermore, the highest recall increased from 81.15% to 96.49%, and the highest F1-score increased to 97.39% from 81.75%. The highest AUC increased from 85.31% to 96.79%.

These findings support the effectiveness of RC selection in improving the accuracy and precision of classification models in our study. Furthermore, we extracted complementary information by ensuring the inter-relations between the modalities and capturing all interacting features, highlighting the importance of inter-relations in the fusion process. RC selection was applied to select 52 features out of 861, which represents 6% of the total number of features. This reduction in the number of features can save time, making the approach more practical for real-world applications, even with the use of cross-validation, which is typically considered a time-consuming validation method. Note that the accuracy can be further improved by adjusting the number of features and classifiers used. The number of selected features was determined to be optimal for the final set.

In *S*MPS experimentation, we applied feature selection after fusing SSMI with MP to investigate the influence of SSMI on the prediction performance. In cross-validation, we observed a significant enhancement in the score of classification. The classification of *S*MPS in corss-validation reached the highest value, at 98.94% from 96.66% in *S*MP. The precision is also improved to 98.27% in *S*MPS from 98.21% in *S*MP. Also, the Recall rate is improved to reach 100% form 98.24%. Additionally, the F1-score is raised from 97.391% *S*MP to 99.13 % in *S*MPS. Moreover, the AUC improved to reach 98.55% from 96.22%. The improvement in accuracy, precision, Recall, F1-score, and AUC demonstrates the role that the new relation SSMI plays in the fusion process. In train-test-split, the accuracy, precision, specificity, recall, F1-score, and AUC are also improved in the train-test-split. It is worth noting that the classification scores achieved using different metrics showed a remarkable improvement compared to the barebone baseline. Notably, the accuracy was improved to 98.94%, compared to the baseline score of 77.69%.

Although RC did not select any PET features during the feature selection process, combining PET and MRI features improved the performance of the AD classification models. In a separate experiment, using RC on MRI data alone resulted in an accuracy of 95.00%, whereas the accuracy was improved to 96.66% when RC was applied to both MRI and PET data. This improvement underscores the importance of combining modalities to extract complementary information. Additionally, even though only two features were selected from the SSMI data, their inclusion still improved the overall performance of the model, highlighting the potential of SSMI in fusion processes even with two features. The various experiments conducted in this study demonstrate the effectiveness of all the modalities and the SSMI in the fusion

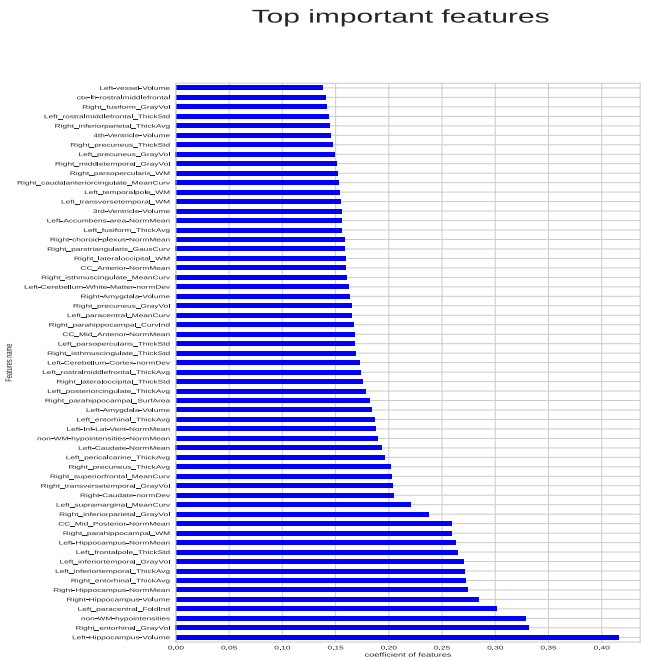


FIGURE 9. Importance of the features listed in Table 19 for AD.vs CN task.

process, emphasizing the importance of respecting specific criteria, such as preserving the local structure information and ensuring intra- and inter-relations between features. These findings provide valuable insights for future research and the application of multimodality fusion techniques.

A. DISCUSSION PREDICTION RESULTS OF VALIDATION SET

In addition to evaluating the proposed method using train-test-split and cross-validation, it is crucial to test its generalizing capability on unseen data, particularly data captured with different protocols and conditions. To achieve this, we used only the features selected by RC from VDR as input, which may serve as biomarkers for new datasets. In VDR train-test-split experiments, our method achieved a significant improvement in accuracy from 88.23% to 100.00% using the biomarkers. In cross-validation, our method achieved the best accuracy of 86.25% using RC-selected features, and this accuracy increased to 98.75% using the biomarkers. These results indicate the robustness and generalizability of our framework, even when applied to new datasets. The AUC values also increased from 80.00% to 96.88% using the biomarkers in the train-test-split experiments. The high precision, recall, and F1-score values obtained by LR, LDA, SVM, and Xgboost were already present in the original raw data, and using the biomarkers did not affect these values significantly. However, the small validation set size and the quality of the MRI and PET used in VDR and the preprocessing may have contributed to the high values obtained.

B. DISCUSSION SELECTED REGIONS

As shown in Table 19, RC selected the *right parahippocampal* region as a biomarker, which has also been identified

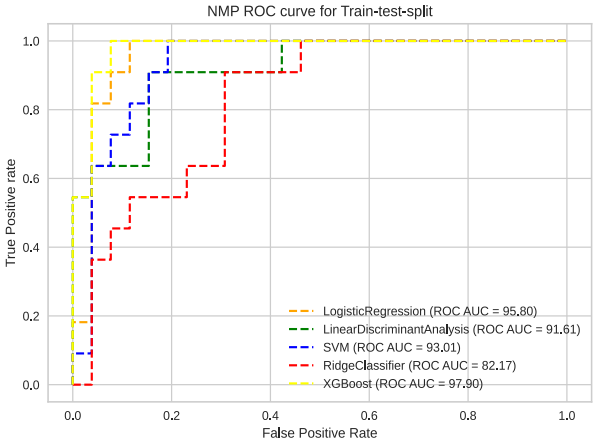


FIGURE 10. NMP ROC curve for each model in train-test-split.

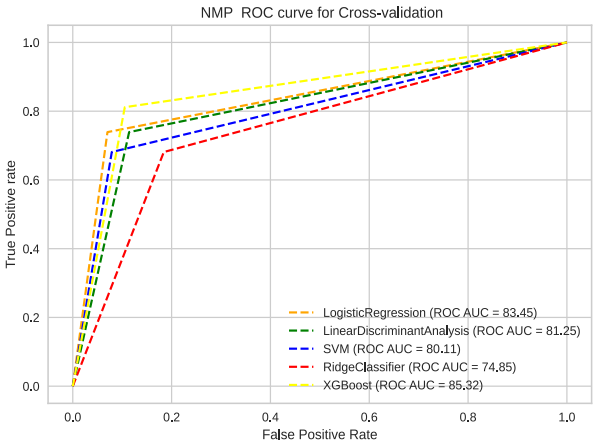


FIGURE 11. NMP ROC curve for each model in cross-validation.

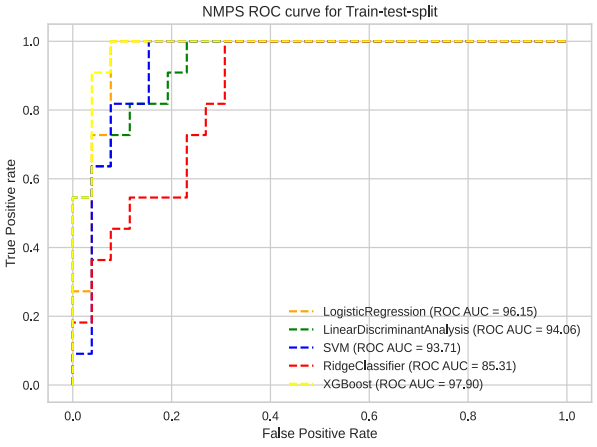


FIGURE 12. NMPS ROC curve for each model in train-test-split.

in [58] as a significant biomarker for distinguishing between AD and CN. The study found significant volume differences between AD and healthy individuals in the parahippocampal gyrus, indicating that its shrinkage could be considered an early sign of AD. The same study also identified the right superior frontal region as a potential biomarker for AD.

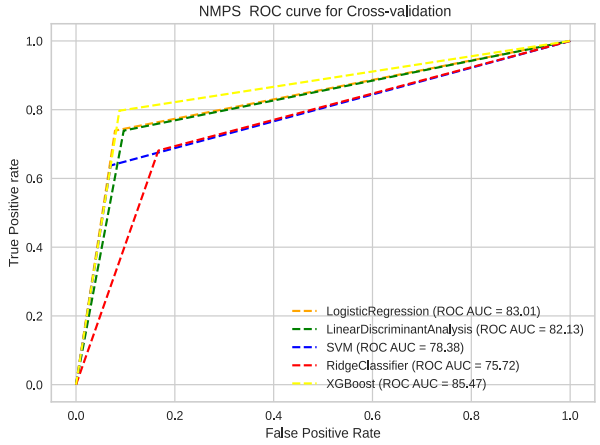


FIGURE 13. NMPS ROC curve for each model in cross-validation.

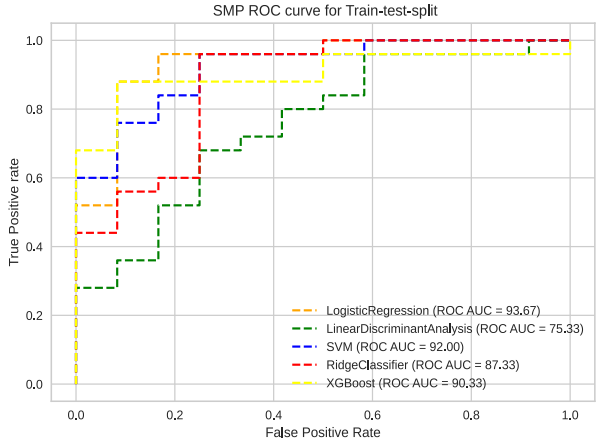


FIGURE 14. SMP ROC curve for each model in train-test-split.

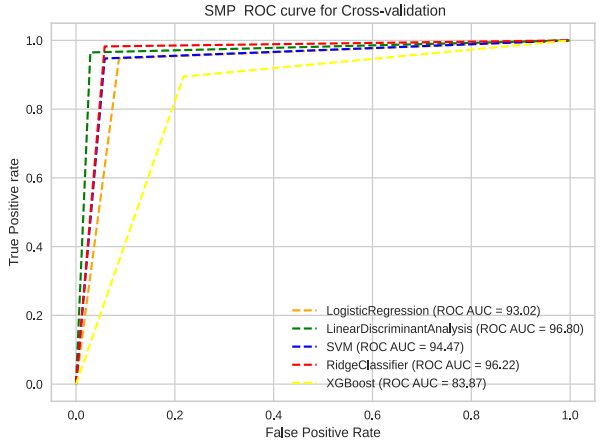


FIGURE 15. SMP ROC curve for each model in cross-validation.

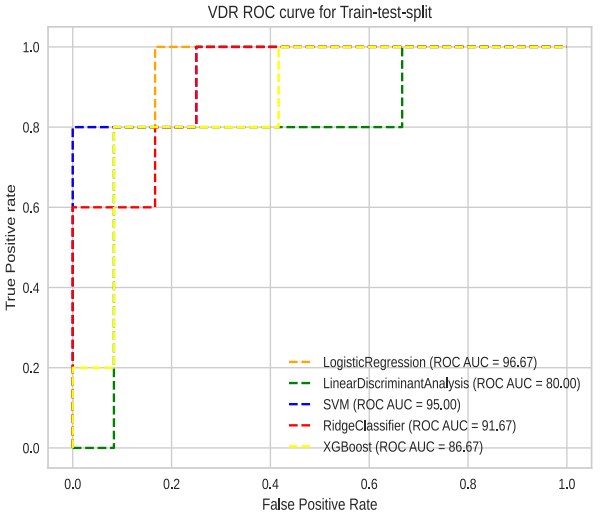


FIGURE 16. VDR ROC curve for each model in train-test-split.

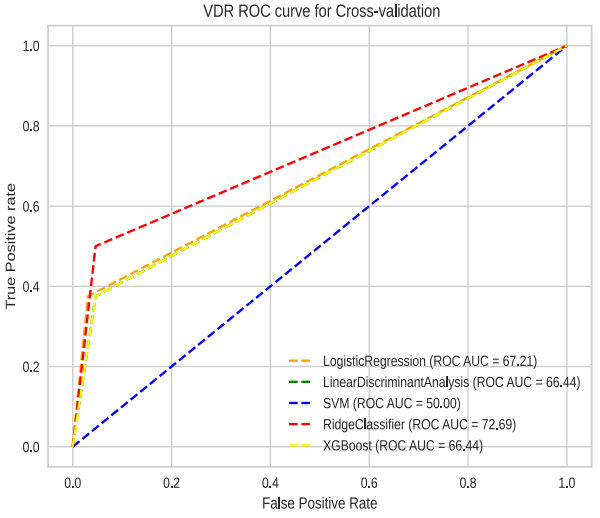


FIGURE 17. VDR-cross ROC curve for each model in cross-validation.

According to [59], all cingulate regions, namely the rostral anterior cingulate, caudal anterior cingulate, and posterior cingulate, were significantly smaller in AD cases compared to controls. The study reported that, considering these cingulate regions' functional and connectional distinctions, identification and monitoring of their atrophy may provide insights into the natural history of AD and may help in the quest for diagnostic markers for early AD. It is noteworthy that RC selected all of these regions.

Bankssts was found to be the most (Aβ)-affected cortical region in both (Aβ)$-$ and (Aβ)$+$ CN participants, and it appears to be more sensitive to identifying early (Aβ) deposition than other regions [60]. Notably, the accumulation of (Aβ) is a symptom of AD. A study by [61] reported that all AD patients showed atrophy in the temporal pole, which may be used as biomarkers for AD classification. [62] examined previous studies that used the parietal lobe to predict AD and concluded that it is clearly involved in the early stages of the disease. Two studies followed patients over the years [63], [64], with the first indicating early atrophy in the left pars opercularis before the development of symptoms. In [65], greater atrophy of various frontal ROIs was associated with increasing agitation and aggressive symptoms: left rostral middle frontal in MCI-stable, right superior frontal in AD.

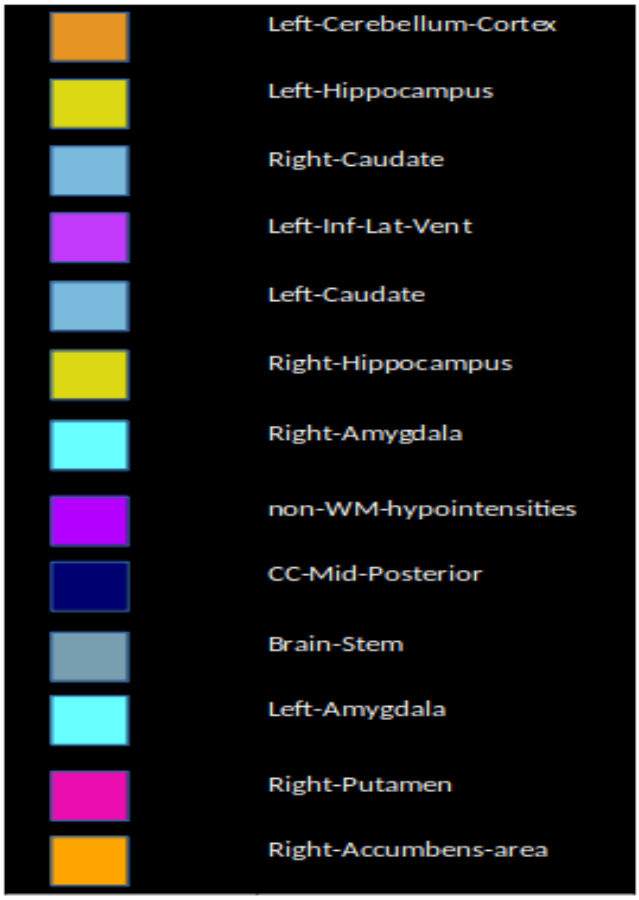


FIGURE 18. Subcortical regions segmentation labels by Freesurfer.

FIGURE 19. Cortical regions segmentation labels by Freesurfer.

Bankssts, parietal lobe, rostral middle frontal, and pars opercularis are selected by RC as discriminative features. The study by [66] found a significant loss of cortical thickness in the right entorhinal cortex, which characterized Alzheimer's patients. In fact, RC chose the right parahippocampal with three measures, in addition to selecting right superior frontal and right entorhinal with two measures, demonstrating its ability to extract discriminative regions related to AD vs. CN classification.

The study by [67] examined whether localized abnormalities of the ventricles may be detected when comparing AD patients to controls, and found that regions surrounding the amygdala, thalamus, tail of the caudate nuclei (particularly the left one), and head of the left caudate nucleus were all considerably impaired in patients with AD. [68] reported that WMH is more common and severe in AD patients than in non-demented individuals. In a longitudinal study, [69] found that the rate of total lateral ventricle enlargement (cm3/yr) differed significantly between Alzheimer's patients and healthy controls, and it was more specific and sensitive to Alzheimer's patients' diagnosis than a comparison of cross-sectional volumes at the final examination. [70] discovered CC atrophy in AD sufferers, and [71] found that the DRN displayed an accumulation of NFTs before the transentorhinal area. [72] noted smaller volumes of the Hippocampus, putamen, and accumbens in AD patients. The Hippocampus is invariably damaged in severe forms of AD, so much so that the disease has been dubbed "hippocampal dementia" [73]. Reference [74] reported a loss of 8% of subjects' Hippocampus before AD symptoms and proposed that Hippocampus atrophy may be used as a hallmark of AD.

In the feature selection step, RC selected the caudate, amygdala, non-WM-hypointensities, lateral ventricle, cc-Mid-Posterior, brain-Stem, and accumbens as discriminative regions in AD classification. The Hippocampus was also selected as a discriminative region, which is consistent with its importance in AD pathology. Figure 9 demonstrates the significance of the chosen feature.

VI. LIMITATION

Our study yielded promising results, with higher accuracy, precision, recall, and F1-score than state-of-the-art methods. However, our framework has some limitations. Firstly, we only utilized two modalities, MRI and PET, while there are other potential modalities, including SPECT or CSF, which could improve classification accuracy. Additionally, demographic information, genetic data, and APOE could

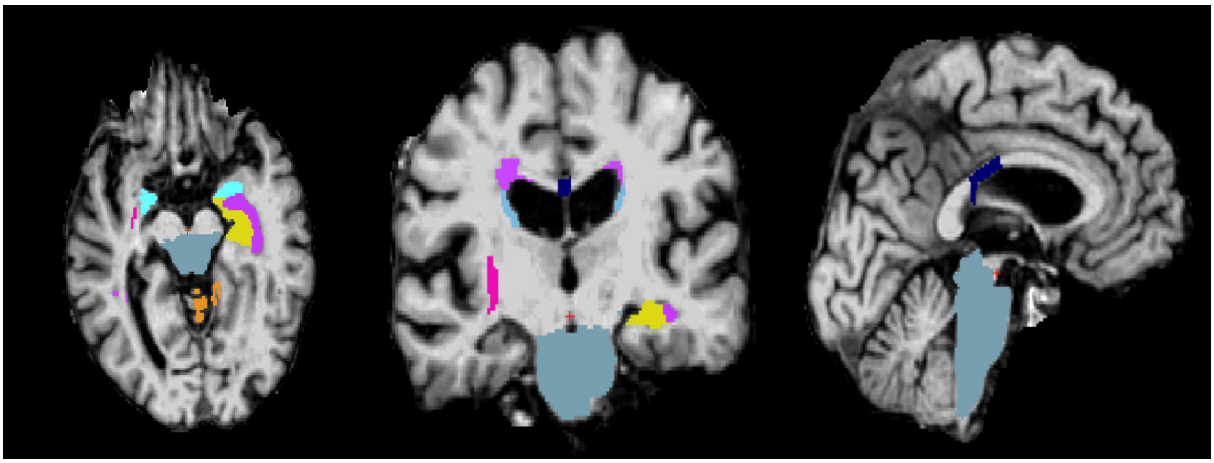


FIGURE 20. Subcortical regions selected by RC.

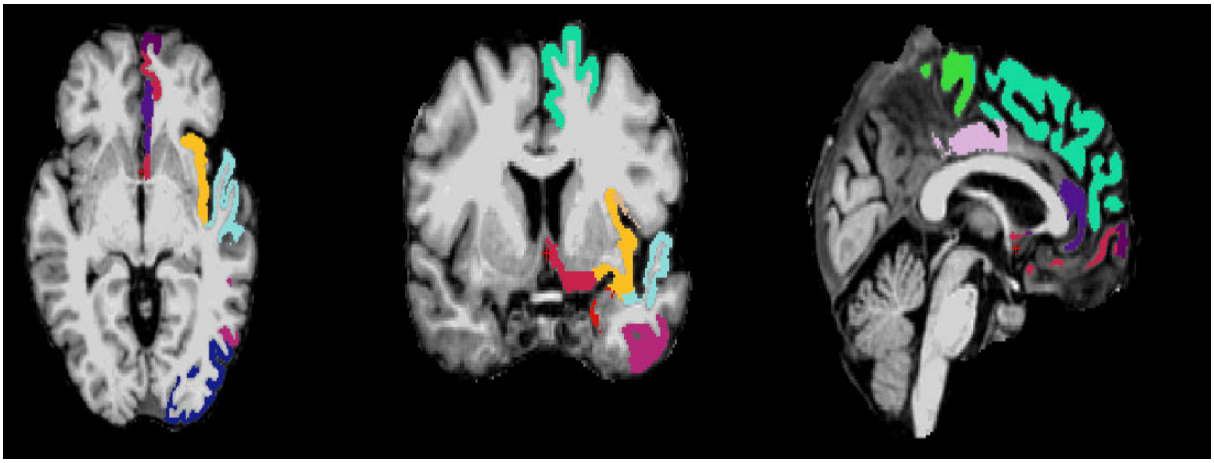


FIGURE 21. Left cortex regions selected by RC.

enhance the learning performance and augment classification results.

Moreover, segmenting PET separately instead of based on MRI segmentation could potentially enhance the classification results. Furthermore, combining MRI and PET sets would impact the intra-relations and should be handled separately for each data source. The parameters were manually selected in terms of validation methods, and applying hyper-parameter tuning methods, such as Gridsearch, during the training step could improve the final classification results.

Additionally, it should be noted that the dataset used in this study was relatively small and mildly imbalanced. Although we used precision, specificity, recall, F1-score, and AUC to evaluate the performance of the learning models, extra work is required to handle the small size and imbalanced nature of the data.

VII. CONCLUSION

In this paper, we present a multimodality fusion approach that combines MRI and PET using machine learning for AD classification. Our main objective was to design a robust fusion framework that adheres to three proposed criteria. We introduced a new relation, SSMI, which captures the interactions between the measures of a specific region for the same patient captured by MRI and PET simultaneously. The approach evaluates the influence of different criteria: intra-relations, SSMI, and inter-relations. Combining the SSMI set with the MRI and PET sets improved the learning performance, and using RC to model the inter-relations further improved the classification accuracy.

We validated our framework using both train-test-split and cross-validation and evaluated its performance using various metrics such as accuracy, precision, specificity, recall, F1-score, and AUC. The results showed significant improvement, demonstrating the robustness of our approach. Our study is unique in that it considers all three relations simultaneously in the same fusion framework and was conducted on a larger sample size compared to previous studies, suggesting the ability to generalize our approach to other classification tasks.

In future work, we plan to expand our approach by adding gene expression data to MRI and PET and incorporating demographic information to improve learning performance. We also aim to enhance the intra-relations by modeling each data source with a non-linear graph and extract more information from PET by increasing the measures of

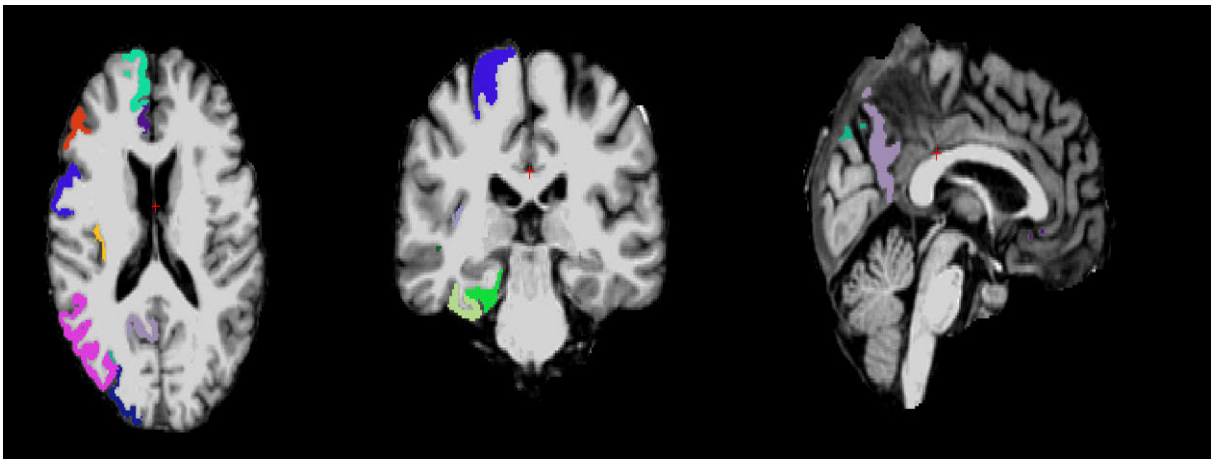


FIGURE 22. Right cortex regions selected by RC.

different regions. Additionally, we propose addressing multi-class classification tasks to classify AD patients and their early stages using SSMI as an exciting research direction. Furthermore, it is crucial to address the issue of imbalanced data in medical imaging by proposing a new approach to oversample the minority class. It is important to consider the sensitivity of medical imaging in the oversampling process to avoid creating artificial patterns that could compromise the accuracy and reliability of the classification model.

APPENDIX A
FIGURES

Figure 10 and Figure 11 show the models' ROC curves in train-test-split and cross-validation, respectively.

The ROC curves for each model in train-test-split and cross-validation for MPS are depicted in Figure 12 and Figure 13, respectively.

Figure 14, and Figure 15 plot the *S*MP ROC curve for each model in train-test-split and cross-validation, respectively.

Figure 16 and Figure 17 plot the VDR ROC curve for each model in train-test-split and cross-validation, respectively.

Figures 18 and Figures 19 show Freesurfer's subcortical and cortical segmentation labels, respectively, while Figures 20, 21, and 22 depict the regions chosen by RC from subcortical, left cortex, and right cortex segmentations, respectively.

ACKNOWLEDGMENT

The authors would like to thank the invaluable assistance of Dr. Belhimer Nadia, who was the radiologue that helped them to verify the quality of MRI preprocessing. Data collection and sharing for this project was funded by the AD Neuroimaging Initiative (ADNI). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd., and its affiliated company Genentech Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research and Development LLC.; Johnson and Johnson Pharmaceutical Research and Development LLC.; Lumosity; Lundbeck; Merck and Company Inc.; Meso Scale Diagnostics LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health. The grantee organization is the Northern California Institute for Research and Education and the study is coordinated by the Alzheimer's Therapeutic Research Institute, University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging, University of Southern California.

REFERENCES

[1] K. A. Johnson, N. C. Fox, R. A. Sperling, and W. E. Klunk, "Brain imaging in Alzheimer disease," *Cold Spring Harb. Perspect. Med.*, vol. 2, no. 4, 2012, Art. no. a006213.

[2] R. R. Raylman, S. Majewski, S. K. Lemieux, S. S. Velan, B. Kross, V. Popov, M. F. Smith, A. G. Weisenberger, C. Zorn, and G. D. Marano, "Simultaneous MRI and PET imaging of a rat brain," *Phys. Med. Biol.*, vol. 51, no. 24, pp. 6371–6379, Dec. 2006.

[3] D. Lahat, T. Adali, and C. Jutten, "Multimodal data fusion: An overview of methods, challenges, and prospects," *Proc. IEEE*, vol. 103, no. 9, pp. 1449–1477, Sep. 2015.

[4] K. Uludağ and A. Roebroeck, "General overview on the merits of multimodal neuroimaging data fusion," *NeuroImage*, vol. 102, pp. 3–10, Nov. 2014.

[5] J. Zhang, "Multi-source remote sensing data fusion: Status and trends," *Int. J. Image Data Fusion*, vol. 1, no. 1, pp. 5–24, Mar. 2010.

[6] T. Tong, K. Gray, Q. Gao, L. Chen, and D. Rueckert, "Multimodal classification of Alzheimer's disease using nonlinear graph fusion," *Pattern Recognit.*, vol. 63, pp. 171–181, Mar. 2017, doi: 10.1016/j.patcog.2016.10.009.

[7] T. Zhang and M. Shi, "Multi-modal neuroimaging feature fusion for diagnosis of Alzheimer's disease," *J. Neurosci. Methods*, vol. 341, Jul. 2020, Art. no. 108795, doi: 10.1016/j.jneumeth.2020.108795.

[8] C. Zu, Y. Wang, L. Zhou, L. Wang, and D. Zhang, "Multi-modality feature selection with adaptive similarity learning for classification of Alzheimer's disease," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 1542–1545.

[9] H. Khachnaoui, M. Agrébi, S. Halouani, and N. Khlifa, "Deep learning for automatic pulmonary embolism identification using CTA images," in *Proc. 6th Int. Conf. Adv. Technol. Signal Image Process. (ATSIP)*, May 2022, pp. 1–6.

[10] H. Khachnaoui, N. Khlifa, and R. Mabrouk, "Machine learning for early Parkinson's disease identification within SWEDD group using clinical and DaTSCAN SPECT imaging features," *J. Imag.*, vol. 8, no. 4, p. 97, Apr. 2022.

[11] Q. Cheng, Q. Zhang, P. Fu, C. Tu, and S. Li, "A survey and analysis on automatic image annotation," *Pattern Recognit.*, vol. 79, pp. 242–259, Jul. 2018, doi: 10.1016/j.patcog.2018.02.017.

[12] V. D. Calhoun and J. Sui, "Multimodal fusion of brain imaging data: A key to finding the missing link(s) in complex mental illness," *Biol. Psychiatry, Cognit. Neurosci. Neuroimag.*, vol. 1, no. 3, pp. 230–244, May 2016, doi: 10.1016/j.bpsc.2015.12.005.

[13] M. Zitnik, F. Nguyen, B. Wang, J. Leskovec, A. Goldenberg, and M. M. Hoffman, "Machine learning for integrating data in biology and medicine: Principles, practice, and opportunities," *Inf. Fusion*, vol. 50, pp. 71–91, Oct. 2019.

[14] F. Wang, L. P. Casalino, and D. Khullar, "Deep learning in medicine-promise, progress, and challenges," *JAMA Internal Med.*, vol. 179, no. 3, pp. 293–294, 2019.

[15] Y. Zhang, S. Wang, and Z. Dong, "Classification of Alzheimer disease based on structural magnetic resonance imaging by kernel support vector machine decision tree," *Prog. Electromagn. Res.*, vol. 144, pp. 171–184, 2014.

[16] M. Liu, D. Cheng, and W. Yan, "Classification of Alzheimer's disease by combination of convolutional and recurrent neural networks using FDG-PET images," *Frontiers Neuroinform.*, vol. 12, p. 35, Jun. 2018.

[17] X. Hao, Y. Bao, Y. Guo, M. Yu, D. Zhang, S. L. Risacher, A. J. Saykin, X. Yao, and L. Shen, "Multi-modal neuroimaging feature selection with consistent metric constraint for diagnosis of Alzheimer's disease," *Med. Image Anal.*, vol. 60, Feb. 2020, Art. no. 101625, doi: 10.1016/j.media.2019.101625.

[18] K. B. Walhovd, A. M. Fjell, J. Brewer, L. K. McEvoy, C. Fennema-Notestine, D. J. Hagler, R. G. Jennings, D. Karow, and A. M. Dale, "Combining MR imaging, positron-emission tomography, and CSF biomarkers in the diagnosis and prognosis of Alzheimer disease," *Amer. J. Neuroradiol.*, vol. 31, no. 2, pp. 347–354, Feb. 2010.

[19] J. Dukart, K. Mueller, H. Barthel, A. Villringer, O. Sabri, and M. L. Schroeter, "Meta-analysis based SVM classification enables accurate detection of Alzheimer's disease across different clinical centers using FDG-PET and MRI," *Psychiatry Res., Neuroimag.*, vol. 212, no. 3, pp. 230–236, Jun. 2013.

[20] P. Forouzannezhad, A. Abbaspour, C. Li, C. Fang, U. Williams, M. Cabrerizo, A. Barreto, J. Andrian, N. Rishe, R. E. Curiel, D. Loewenstein, R. Duara, and M. Adjouadi, "A Gaussian-based model for early detection of mild cognitive impairment using multimodal neuroimaging," *J. Neurosci. Methods*, vol. 333, Mar. 2020, Art. no. 108544.

[21] D. Zhang, Y. Wang, L. Zhou, H. Yuan, and D. Shen, "Multimodal classification of Alzheimer's disease and mild cognitive impairment," *NeuroImage*, vol. 55, no. 3, pp. 856–867, Apr. 2011.

[22] J. Young, M. Modat, M. J. Cardoso, A. Mendelson, D. Cash, and S. Ourselin, "Accurate multimodal probabilistic prediction of conversion to Alzheimer's disease in patients with mild cognitive impairment," *NeuroImage, Clin.*, vol. 2, pp. 735–745, Jan. 2013.

[23] D. Zhang and D. Shen, "Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease," *NeuroImage*, vol. 59, no. 2, pp. 895–907, Jan. 2012.

[24] F. Liu, C.-Y. Wee, H. Chen, and D. Shen, "Inter-modality relationship constrained multi-modality multi-task feature selection for Alzheimer's disease and mild cognitive impairment identification," *NeuroImage*, vol. 84, pp. 466–475, Jan. 2014, doi: 10.1016/j.neuroimage.2013.09.015.

[25] C. Zu, B. Jie, M. Liu, S. Chen, D. Shen, and D. Zhang, "Label-aligned multi-task feature learning for multimodal classification of Alzheimer's disease and mild cognitive impairment," *Brain Imag. Behav.*, vol. 10, no. 4, pp. 1148–1159, Dec. 2016.

[26] A. Ortiz, D. Fajardo, J. M. Górriz, J. Ramírez, and F. J. Martínez-Murcia, "Multimodal image data fusion for Alzheimer's Disease diagnosis by sparse representation," *Stud. Health Technol. Inform.*, vol. 207, pp. 11–18, 2014.

[27] L. Xu, X. Wu, K. Chen, and L. Yao, "Multi-modality sparse representation-based classification for Alzheimer's disease and mild cognitive impairment," *Comput. Methods Programs Biomed.*, vol. 122, no. 2, pp. 182–190, Nov. 2015, doi: 10.1016/j.cmpb.2015.08.004.

[28] V. Youssofzadeh, B. McGuinness, L. P. Maguire, and K. Wong-Lin, "Multi-kernel learning with Dartel improves combined MRI-PET classification of Alzheimer's disease in AIBL data: Group and individual analyses," *Frontiers Hum. Neurosci.*, vol. 11, pp. 1–12, Jul. 2017.

[29] B. Fischl, "FreeSurfer," *NeuroImage*, vol. 62, no. 2, pp. 774–781, Aug. 2012.

[30] S. Sargolzaei, A. Sargolzaei, M. Cabrerizo, G. Chen, M. Goryawala, S. Noei, Q. Zhou, R. Duara, W. Barker, and M. Adjouadi, "A practical guideline for intracranial volume estimation in patients with Alzheimer's disease," *BMC Bioinf.*, vol. 16, no. S7, pp. 1–10, Dec. 2015.

[31] S. D. Edland, Y. Xu, M. Plevak, P. O'Brien, E. G. Tangalos, R. C. Petersen, and C. R. Jack, "Total intracranial volume: Normative values and lack of association with Alzheimer's disease," *Neurology*, vol. 59, no. 2, pp. 272–274, Jul. 2002.

[32] J. A. Thie, "Understanding the standardized uptake value, its methods, and implications for usage," *J. Nucl. Med.*, vol. 45, no. 9, pp. 1431–1434, 2004.

[33] C. Li, C. Fang, M. Cabrerizo, A. Barreto, J. Andrian, R. Duara, D. Loewenstein, and M. Adjouadi, "Pattern analysis of the interaction of regional amyloid load, cortical thickness and APOE genotype in the progression of Alzheimer's disease," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Nov. 2017, pp. 2171–2176.

[34] C. Fang, C. Li, P. Forouzannezhad, M. Cabrerizo, R. E. Curiel, D. Loewenstein, R. Duara, and M. Adjouadi, "Gaussian discriminative component analysis for early detection of Alzheimer's disease: A supervised dimensionality reduction algorithm," *J. Neurosci. Methods*, vol. 344, Oct. 2020, Art. no. 108856, doi: 10.1016/j.jneumeth.2020.108856.

[35] W. Shao, L. He, C.-T. Lu, X. Wei, and P. S. Yu, "Online unsupervised multi-view feature selection," in *Proc. IEEE 16th Int. Conf. Data Mining (ICDM)*, Dec. 2016, pp. 1203–1208.

[36] J. Li, L. Wu, G. Wen, and Z. Li, "Exclusive feature selection and multi-view learning for Alzheimer's disease," *J. Vis. Commun. Image Represent.*, vol. 64, Oct. 2019, Art. no. 102605, doi: 10.1016/j.jvcir.2019.102605.

[37] P. A. Freeborough, N. C. Fox, and R. I. Kitney, "Interactive algorithms for the segmentation and quantitation of 3-D MRI brain scans," *Comput. Methods Programs Biomed.*, vol. 53, no. 1, pp. 15–25, May 1997.

[38] H. Abdi and L. J. Williams, "Principal component analysis," *WIREs Comput. Statistic*, vol. 2, no. 4, pp. 433–459, Jul./Aug. 2010.

[39] J. Miao and L. Niu, "A survey on feature selection," *Proc. Comput. Sci.*, vol. 91, pp. 919–926, Jan. 2016.

[40] G. C. McDonald, "Ridge regression," *Wiley Interdiscipl. Rev. Comput. Stat.*, vol. 1, no. 1, pp. 93–100, 2009.

[41] I. Naseem, R. Togneri, and M. Bennamoun, "Linear regression for face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 11, pp. 2106–2112, Nov. 2010.

[42] S. Menard, *Applied Logistic Regression Analysis*, no. 106. Newbury Park, CA, USA: Sage, 2002.

[43] A. Tharwat, "Linear vs. quadratic discriminant analysis classifier: A tutorial," *Int. J. Appl. Pattern Recognit.*, vol. 3, no. 2, pp. 145–180, 2016.

[44] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intell. Syst. Appl.*, vol. 13, no. 4, pp. 18–28, Jul. 1998.

[45] O. Sagi and L. Rokach, "Ensemble learning: A survey," *WIREs Data Mining Knowl. Discovery*, vol. 8, no. 4, Jul. 2018, Art. no. e1249.

[46] P. Refaeilzadeh, L. Tang, H. Liu, L. Angeles, and C. D. Scientist, "Cross-validation," *Encyclopedia Database Syst.*, vol. 5, pp. 532–538, Jan. 2020.

[47] F. Gorunescu, "Classification performance evaluation," in *Data Mining*. Cham, Switzerland: Springer, 2011, pp. 319–330.

[48] N. Japkowicz, "The class imbalance problem: Significance and strategies," in *Proc. Int. Conf. Artif. Intell.*, vol. 56, 2000, pp. 111–117.

[49] O. Kohannim, X. Hua, D. P. Hibar, S. Lee, Y.-Y. Chou, A. W. Toga, C. R. Jack, M. W. Weiner, and P. M. Thompson, "Boosting power for clinical trials using classifiers based on multiple biomarkers," *Neurobiol. Aging*, vol. 31, no. 8, pp. 1429–1442, Aug. 2010, doi: 10.1016/j.neurobiolaging.2010.04.022.

[50] Y. Wang, M. Liu, L. Guo, and D. Shen, "Kernel-based multi-task joint sparse classification for Alzheimer'S disease," in *Proc. IEEE 10th Int. Symp. Biomed. Imag.*, Apr. 2013, pp. 1364–1367.

[51] Y. Zhang, S. Wang, K. Xia, Y. Jiang, and P. Qian, "Alzheimer's disease multiclass diagnosis via multimodal neuroimaging embedding feature selection and fusion," *Inf. Fusion*, vol. 66, pp. 170–183, Feb. 2021.

[52] G. B. Frisoni, N. C. Fox, C. R. Jack, P. Scheltens, and P. M. Thompson, "The clinical use of structural MRI in Alzheimer disease," *Nature Rev. Neurol.*, vol. 6, no. 2, pp. 67–77, Feb. 2010.

[53] M. P. van den Heuvel and H. E. Hulshoff Pol, "Exploring the brain network: A review on resting-state fMRI functional connectivity," *Eur. Neuropsychopharmacol.*, vol. 20, no. 8, pp. 519–534, Aug. 2010.

[54] A. Vinayagam, J. Zirin, C. Roesel, Y. Hu, B. Yilmazel, A. A. Samsonova, R. A. Neumüller, S. E. Mohr, and N. Perrimon, "Erratum: Integrating protein–protein interaction networks with phenotypes reveals signs of interactions," *Nature Methods*, vol. 11, no. 7, pp. 94–99, Jul. 2014.

[55] K. A. Johnson et al., "Imaging of amyloid burden and distribution in cerebral amyloid angiopathy," *Ann. Neurology, Off. J. Amer. Neurolog. Assoc. Child Neurol. Soc.*, vol. 62, no. 3, pp. 229–234, 2007.

[56] C. R. Jack, "Suspected non-alzheimer disease pathophysiology-concept and controversy," *Nature Rev. Neurol.*, vol. 12, no. 2, pp. 117–124, 2016.

[57] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets," *PLoS ONE*, vol. 10, no. 3, Mar. 2015, Art. no. e0118432.

[58] C. Echávarri, P. Aalten, H. B. M. Uylings, H. I. L. Jacobs, P. J. Visser, E. H. B. M. Gronenschild, F. R. J. Verhey, and S. Burgmans, "Atrophy in the parahippocampal gyrus as an early biomarker of Alzheimer's disease," *Brain Struct. Function*, vol. 215, nos. 3–4, pp. 265–271, Jan. 2011.

[59] B. F. Jones, J. Barnes, H. B. M. Uylings, N. C. Fox, C. Frost, M. P. Witter, and P. Scheltens, "Differential regional atrophy of the cingulate gyrus in Alzheimer disease: A volumetric MRI study," *Cerebral Cortex*, vol. 16, no. 12, pp. 1701–1708, Dec. 2005.

[60] T. Guo, S. M. Landau, and W. J. Jagust, "Detecting earlier stages of amyloid deposition using PET in cognitively normal elderly adults," *Neurology*, vol. 94, no. 14, pp. e1512–e1524, Apr. 2020.

[61] S. E. Arnold, B. T. Hyman, and G. W. Van Hoesen, "Neuropathologic changes of the temporal pole in Alzheimer's disease and pick's disease," *Arch. Neurol.*, vol. 51, no. 2, pp. 145–150, Feb. 1994.

[62] H. I. L. Jacobs, M. P. J. Van Boxtel, J. Jolles, F. R. J. Verhey, and H. B. M. Uylings, "Parietal cortex matters in Alzheimer's disease: An overview of structural, functional and metabolic findings," *Neurosci. Biobehav. Rev.*, vol. 36, no. 1, pp. 297–309, Jan. 2012.

[63] M. L. Gorno-Tempini, R. C. Murray, K. P. Rankin, M. W. Weiner, and B. L. Miller, "Clinical, cognitive and anatomical evolution from nonfluent progressive aphasia to corticobasal syndrome: A case report," *Neurocase*, vol. 10, no. 6, pp. 426–436, Dec. 2004.

[64] J. C. Janssen, "Mapping the onset and progression of atrophy in familial frontotemporal lobar degeneration," *J. Neurol., Neurosurg. Psychiatry*, vol. 76, no. 2, pp. 162–168, Feb. 2005.

[65] P. T. Trzepacz, P. Yu, P. K. Bhamidipati, B. Willis, T. Forrester, L. Tabas, A. J. Schwarz, and A. J. Saykin, "Frontolimbic atrophy is associated with agitation and aggression in mild cognitive impairment and Alzheimer's disease," *Alzheimer's Dementia*, vol. 9, no. 5S, pp. S95–S104, Oct. 2013.

[66] H. Yang, H. Xu, Q. Li, Y. Jin, W. Jiang, J. Wang, Y. Wu, W. Li, C. Yang, X. Li, S. Xiao, F. Shi, and T. Wang, "Study of brain morphology change in Alzheimer's disease and amnestic mild cognitive impairment compared with normal controls," *Gen. Psychiatry*, vol. 32, no. 2, Apr. 2019, Art. no. e100005.

[67] L. Ferrarini, W. M. Palm, H. Olofsen, M. A. van Buchem, J. H. C. Reiber, and F. Admiraal-Behloul, "Shape differences of the brain ventricles in Alzheimer's disease," *NeuroImage*, vol. 32, no. 3, pp. 1060–1069, Sep. 2006.

[68] A. M. Brickman, J. Muraskin, and M. E. Zimmerman, "Structural neuroimaging in Alzheimer's disease: Do white matter hyperintensities matter?" *Dialogues Clin. Neurosci.*, vol. 11, no. 2, pp. 181–190, Jun. 2009.

[69] C. D. Md, J. V. Haxby, J. A. Gillette, D. Teichberg, S. I. Rapoport, and M. B. Schapiro, "Longitudinal changes in lateral ventricular volume in datients with dementia of the Alzheimer type," *Neurology*, vol. 42, no. 10, p. 2029, Oct. 1992.

[70] M. Zhu, X. Wang, W. Gao, C. Shi, H. Ge, H. Shen, and Z. Lin, "Corpus callosum atrophy and cognitive decline in early Alzheimer's disease: Longitudinal MRI study," *Dementia Geriatric Cognit. Disorders*, vol. 37, nos. 3–4, pp. 214–222, 2014.

[71] L. T. Grinberg, U. Rüb, R. E. L. Ferretti, R. Nitrini, J. M. Farfel, L. Polichiso, K. Gierga, W. Jacob-Filho, and H. Heinsen, "The dorsal raphe nucleus shows phospho-tau neurofibrillary changes before the transentorhinal region in Alzheimer's disease. A precocious onset?" *Neuropathol. Appl. Neurobiol.*, vol. 35, no. 4, pp. 406–416, Aug. 2009.

[72] B. Patenaude, S. M. Smith, D. N. Kennedy, and M. Jenkinson, "A Bayesian model of shape and appearance for subcortical brain segmentation," *NeuroImage*, vol. 56, no. 3, pp. 907–922, Jun. 2011.

[73] B. T. Hyman, G. W. Van Hoesen, A. R. Damasio, and C. L. Barnes, "Alzheimer's disease: Cell-specific pathology isolates the hippocampal formation," *Science*, vol. 225, no. 4667, pp. 1168–1170, Sep. 1984.

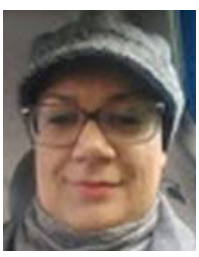
[74] N. C. Fox, E. K. Warrington, P. A. Freeborough, P. Hartikainen, A. M. Kennedy, J. M. Stevens, and M. N. Rossor, "Presymptomatic hippocampal atrophy in Alzheimer's disease: A longitudinal MRI study," *Brain*, vol. 119, no. 6, pp. 2001–2007, 1996.

BOUCHRA GUELIB received the bachelor's degree from the Faculty of Nouvelle Téchnologies d'Informations et Communication, University of Abdelhamid Mehri Constantine 2, in 2015, and the M.S. degree in information systems from the University of Abdelhamid Mehri Constantine 2, in 2018, where she is currently pursuing the Ph.D. degree with the LIRE Laboratory, Department of TLSI, with a focus on multimodal fusion using machine learning. Her research interests include medical image processing, multimodal fusion, and bioinformatic field.

KARIM ZAROUR received the Ph.D. degree from the Mentouri University of Constantine and the Habilitation degree from the University of Abdelhamid Mehri Constantine 2, Algeria. He is currently an Associate Professor in computer science with the University of Abdelhamid Mehri Constantine 2. He has supervises many Ph.D. and master's students. He has published many articles in international journals and conferences. His current research interests include health informatics, IA, privacy and security in healthcare, and multiagent systems and cloud.

HAITHEM HERMESSI received the engineering degree in computer science from the Tunisian Aviation School of Borj Amri (Air Force Academy), in 2007, the master's degree from the Higher Institute of Management Sousse, Tunisia, in 2013, and the Ph.D. degree from the LIMTIC Laboratory, University of Tunis El Manar, Tunisia, in 2020. His research interests include with computer vision (generative models, vision transformers, and self-supervised representation learning) and pattern recognition.

BOUNAB RAYENE received the bachelor's degree in computer science from the University of Abdelhamid Mehri Constantine 2, in 2016, and the master's degree from the Faculty of Nouvelle Téchnologies d'Informations et Communication (NTIC), University of Abdelhamid Mehri Constantine 2, in 2019, where she is currently pursuing the Ph.D. degree with the LIRE Laboratory, Department of TLSI. She is also working on machine learning for healthcare fraud detection.



KHLIFA NAWRES received the Ph.D. degree from the National School of Engineers of Tunis (ENIT). She is currently a Professor with the Higher Institute of Medical Technologies of Tunis, Tunis El Manar University. She is also an engineer. She coordinates the TIMEd Team: Medical Image Processing of the BTM Laboratory. Her research interests include artificial intelligence and CAD design in medical imaging, emotion recognition, and gaze tracking.